DOCUMENT RESUME

ED 262 977 SE 046 189

AUTHOR	Grey, D. R., Ed.; And Others
TITLE	Proceedings of the International Conference on Teaching Statistics (1st, University of Sheffield, England, August 9-13, 1982). Volume II.
REPORT NO	ISBN-0-946554-02-1
PUB DATE	83
NOTE	413p.; Edited and published by the Organising Committee of the First International Conference on Teaching Statistics. For Volume I, see SE 046 188.
AVAILABLE FROM	Teaching Statistics Trust, c/o Centre for Statistical Education, 25 Broomgrove Road, Sheffield, United Kingdom, S10 2NA, (10 pounds 50 pence).
PUB TYPE	Collected Works - Conference Proceedings (021)
EDRS PRICE	MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS	Agriculture; *Developing Nations; Elementary Secondary Education; Higher Education; Industry; *Mathematics Education; *Mathematics Instruction; *Probability; *Statistics; *Teaching Methods

ABSTRACT

The second volume of these conference proceedings includes 35 papers presented under seven topic headings. The topic headings are: (1) education and training in statistics in developing countries; (2) co-operation between academic and practicing statisticians; (3) teaching survey sampling; (4) teaching design and analysis of experiments; (5) training industrial practitioners; (6) principles of learning probability and statistics (including papers examining children's choice behavior in probabilistic situations, a survey of probability concepts in 3000 students aged 11-16 years, teaching probability and statistics so as to overcome some misconceptions, and the role of explanation in teaching elementary probability); and (7) the teaching of agricultural statistics (which includes a single paper on training of personnel for agricultural statistics in Jordan). Like the papers in the first volume, these papers provide a summary of the situation in statistical education around the world and point the way to future developments. (JN)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED262977

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)

☑ This document has been reproduced as received from the person or organisation originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

ICOTS

PROCEEDINGS
# of the
FIRST INTERNATIONAL CONFERENCE ON TEACHING STATISTICS

VOLUME II

SE 046 189

"PERMISSION TO REPRODUCE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY
V. Barnett

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

University of Sheffield
9 – 13 August 1982
2

ICOTS

University of Sheffield
9 – 13 August 1982

PROCEEDINGS
# of the
FIRST INTERNATIONAL CONFERENCE ON TEACHING STATISTICS

VOLUME II

Edited by:
Organising Committee of the First International Conference on Teaching Statistics

D R Grey, P Holmes, V Barnett, G M Constable

Published by:
Organising Committee of the First International Conference on Teaching Statistics
D R Grey, P Holmes, V Barnett, G M Constable

© Organising Committee of the First International Conference on Teaching Statistics 1983

ISBN 0 946554 00 5 (set of two volumes)

ISBN 0 946554 01 3 (Volume I)

ISBN 0 946554 02 1 (Volume II)

Printed by the University of Sheffield Printing Unit

Published by Teaching Statistics Trust

"It was hard work . . .

. . . but not all the time".

CONTENTS OF VOLUME I

Page No.

Preface xi

INTRODUCTORY TALK

Why Teach Statistics?
V Barnett 3

PLENARY TALKS

Stochastics at the School Level in the Age of the Computer
L Råde 19

Optimum Balance between Statistical Theory and Applications in Teaching
C Radhakrishna Rao 34

The Relevance of Statistical Training
J Gani 50

TOPIC 1: TEACHING STATISTICS IN SCHOOLS (6-11 YEARS AGE GROUP)
ORGANISER: PROFESSOR K MATUSITA

Statistics in the Curriculum for Everybody - how Young Children and how their Teachers react
Tamas Varga 71

Teaching of Statistics in Elementary Schools in Japan
T Yanagawa 81

Two Examples of Statistics Activities for Elementary School Classes
Clare Heidema 91

Using Games to teach some Probability Concepts
George W Bright, John G Harvey and Margariete M Wheeler 110

Teaching Statistics in Elementary Schools in Japan
H Midzuno 116

TOPIC 2: TEACHING STATISTICS IN SCHOOLS (12-15 YEARS AGE GROUP)
ORGANISER: PROFESSOR P L HENNEQUIN

Why Teach Descriptive Statistics in Secondary Schools?
Heinrich Winter 127

The Teaching of Statistics in French Secondary Schools
G Dumousseau 144

Teaching Statistics Concepts through Instructional Games
George W Bright, John C Harvey and Margariete M Wheeler 158

A Practical Approach to Statistics
R Kapadia 169

On the Teaching of Order Statistics
K Bognár 179

TOPIC 3: TEACHING STATISTICS IN SCHOOLS (16-18 YEARS AGE GROUP)
ORGANISER: PROFESSOR J SWIFT

Amount and Kind of Statistics to be taught to 16+ year olds
F Eicker 191

Teaching Statistics to 16-18 year olds: Dreams and Reality
T Nemetz 199

Statistics for the 16-18 year olds: a Discussion
A Hanley 208

TOPIC 4: TEACHING STATISTICS TO NON-STATISTICIANS
ORGANISER: DR W G GILCHRIST

We must preach what is practised
A S C Ehrenberg 215

Teaching Statistics to Non-Statisticians - an Application oriented Approach
G Sadasivan 219

Utility through Collaboration: a Problem oriented Approach
D Taylor and G E Skipworth 232

An Open University Service Course
John Bibby 238

The Training of Service Teachers
R W Madsen 250

TOPIC 5: DEVELOPMENT OF TEACHING MATERIALS AT SCHOOL LEVEL
ORGANISER: MR P HOLMES

Development of Teaching Materials at School Level
A R Hoffer 263

Developing Materials for the Schools Council Project on Statistical Education
P Holmes 269

EDA for 16-19's?
J Bibby 277

TOPIC 6: TRAINING OF TEACHERS IN STATISTICS
ORGANISER: PROFESSOR A P SHULTE

Preparations for College Teaching in Statistics
A K Gupta 283

Teaching Design of Experiments in Indian Universities
V Bhagwandas 287

Clinical Approach to training Statistics Teachers
J Rustagi 298

Discussion of Teacher-Training Papers
A Morley 302

TOPIC 7: USE OF CALCULATORS AND COMPUTERS IN TEACHING
ORGANISER: PROFESSOR L RÅDE

The Case for using Computers in teaching Statistics
B L Joiner 307

Computer Algorithms for the Classical Distribution Functions
H G Posten 313

Use of hand held Programmable Calculators in Statistics Courses for Nontechnical Students
D R Barr 331

Simulation of Games studied in the Seventeenth and Eighteenth Centuries
J Badrikian 336

TOPIC 8: TEACHING STATISTICS WITH THE HELP OF CASE STUDIES
ORGANISER: PROFESSOR B EPSTEIN

Case Studies for an adequate understanding and interpretation of Results by Statistical Inference
M Borovcnik 341

Is a Scientific Legend a comprehensive Case Study?
V V Shvyrkov and A C Davis 356

TOPIC 9: TRAINING STATISTICAL PRACTITIONERS
ORGANISER: PROFESSOR M BENYAKLEF

Training of Statistical Practitioners
P Osinski 375

Preparing Statisticians for Government Service - a View from the United States
Marie D Eldridge and Katherine K Wallman 389

Programmes d'Enseignement et Couts de Formation
M Benyaklef 403

Les Enseignements Statistiques a l'ENSAE
Ph Tassi et A Trognon 415

CONTENTS OF VOLUME II

Page
No.

TOPIC 10: EDUCATION AND TRAINING IN STATISTICS IN DEVELOPING COUNTRIES
ORGANISER: DR M N MURTHY

Statistical Education and Training in Latin America: a General Appraisal
Enrique Cansado 431

Statistical Education and Training in the Developing Countries in Africa
J N Adichie and Biyi Afonja 439

Statistical Education and Training in Developing Countries of Asia and the Pacific
M N Murthy 450

Comments on the above Papers
K T de Graft-Johnson 483

Report on session on Statistical Education and Training in Developing Countries
B P Adhikari 487

TOPIC 11: CO-OPERATION BETWEEN ACADEMIC AND PRACTISING STATISTICIANS
ORGANISER: PROFESSOR W G HUNTER

Co-operation between University and Industry Statisticians in the United States
R D Snee 491

Some Observations from Europe
Lars-Erik Öller 503

Building Bridges between the Academic and Real Worlds - some Observations from South America
C A Peres and P A Morettin 509

Building Bridges between the Academic and Real Worlds
S Jones, B J T Morgan and G B Wetherill 521

Closing the Gap between Statisticians and Statistics Users
Guido E del Pino and Carlos Prado-Campos 527

Co-operation between Academic and Practising Statisticians
A Prat Bartés 536

TOPIC 12: TEACHING SURVEY SAMPLING
ORGANISER: DR D SINGH

The Teaching of Sample Design to Government Statisticians from Developing Countries
Joseph E Garrett 541

Teaching of Survey Sampling
D Singh 552

Teaching of Survey Sampling
P K Bose 566

TOPIC 13: TEACHING DESIGN AND ANALYSIS OF EXPERIMENTS
ORGANISER: DR D A PREECE

The Design and Analysis of Experiments – what is it about?
D A Preece 575

Teaching Experimental Design in the Computer Age
R Mead 591

Teaching Concepts of Design and Analysis using Real and Simulated Experiments
J R Elliott and J C Robinson 603

Teaching Experimental Design: prescribed by a Medical Statistician
Sheila M Gore 623

Contribution to Discussion of the above Papers
B Jones 633

TOPIC 14: TRAINING INDUSTRIAL PRACTITIONERS
ORGANISER: DR R D SNEE

The Polymath Consultant
A A Greenfield 641

Teaching Applied Statistics for and/or in Industry
H C Hamaker 655

Practicals, Case Studies, Projects and Professional Experience: some Methods of developing Skills among Budding Statisticians
G K Kanji 671

Training in Statistical Quality Control
A F Bissell 685

Discussion of Papers by Greenfield, Hamaker and Kanji
A F Bissell 691

Discussion of Papers by Greenfield, Hamaker and Kanji
D G Watts 695

TOPIC 15: PRINCIPLES OF LEARNING PROBABILITY AND STATISTICS
ORGANISER: PROFESSOR E FISCHBEIN

Learning from Experience – Bayes' Theorem: a Model for Stochastic Learning Situations?
H Steinbring and G von Harten 701

Children's Choice Behaviour in Probabilistic Situations
Ruma Falk 714

An Idea of structuring Probability Teaching based on Dienes's Six Stages
Anna Olecka 727

Does the Teaching of Probability improve Probabilistic Intuitions?
E Fischbein and A Gazit 738

Stochastics Teaching based on Common Sense
H-J Bentz 753

A Survey of Probability Concepts in 3000 Pupils aged 11-16 Years
D R Green 766

Misconceptions of Probability, systematic and otherwise; teaching Probability and Statistics so as to overcome some Misconceptions
J M Shaughnessy 784

The Role of Explanation in Teaching Elementary Probability
J L Myers 802

Probabilities: Learning by doing or learning by imitating
A Abele 813

TOPIC 19: TEACHING OF AGRICULTURAL STATISTICS
ORGANISER: DR D SINGH

Training of Personnel for Agricultural Statistics in Jordan
Z Takrouri 831

PREFACE

The first International Conference on Teaching Statistics (ICOTS) took place in the University of Sheffield over the period 9-13 August, 1982. The Conference attracted much interest from all corners of the world. Over 400 full participants from more than 60 countries (accompanied by 100 or so associate participants) spent an active week attending Plenary Sessions, Invited Talks, Contributed Paper Sessions, Workshops and a variety of social events including walks in the Peak District hills and a reception and buffet generously provided by the Lord Mayor of Sheffield on behalf of the City.

The formal programme was very full and covered a wide range of interests in statistical education and training, including the teaching of statistics for all age groups in schools and colleges, the training of practitioners in different applied disciplines, the special problems in developing countries and the difficulties of teaching specific topics. An opening address was given by Professor S. Gregory, Pro-Vice-Chancellor of the University of Sheffield and the closing session was chaired by Mr. H. Ben-Amor, the Director of the UNESCO Office of Statistics.

The Conference was sponsored by the International Statistical Institute and the organisers are grateful to the ISI, and to UNESCO, for financial support. The technical programme and local organisation were arranged by the Programme Committee and the Organising Committee under the respective chairmanships of Professor L. Råde and Professor V. Barnett. The full membership of the two committees was as follows:

Programme Committee

L. Råde (Chairman)
E. Cansado
P. Holmes
B. L. Joiner
M. N. Murthy
G. E. Noether
J. O. Oyelese
B. Penkov
R. Zielinski

Organising Committee

Professor V. Barnett (Chairman)
G. M. Constable
D. R. Grey
P. Holmes
A. Barnett (Conference Secretary)

This Conference arose from the special interests and actions of the International Statistical Institute (ISI): in particular, from the establishment in 1978 of an Education Committee Task Force charged with organising an international conference on the teaching of statistics. The field of statistical education has become a very extensive one and the need for international conferences on the teaching of statistics has been emphasised on several occasions, especially at the many round table meetings on the teaching of statistics arranged by

ISI, and at the international conferences on mathematical education organised by ICMI (International Commission on Mathematics Instruction).

These published Proceedings of the first ICOTS consist of all plenary and invited papers together with contributions from invited discussants, collected together under appropriate topic headings. It is regrettable that the sheer volume of this material has precluded the possibility of including any of the very worthy contributed papers.

It is appropriate to acknowledge the vast effort extended by all members of the Programme and Organising Committees in setting up the Conference and to thank all Topic Organisers for helping to provide an interesting programme. In addition to thanking all authors for their co-operation in submitting papers in suitable form for publication, it is necessary to express particular indebtedness to three people who have given invaluable assistance in bringing this publication into being: Dr. David Grey, who organised the overall preparation of the Proceedings, Mrs. Jenny Cameron, who has undertaken the mammoth task of re-typing all the papers in standard format, and Mr. Bill Seaton, one of the Conference participants, who during an extended period of sabbatical leave in Sheffield, has willingly devoted many hours to the onerous task of proof-reading.

In accordance with views expressed by Conference participants it is hoped that this Conference will be the first of a series of such international conferences on teaching statistics to be held at four-yearly intervals.

V. Barnett
(Chairman, Organising Committee)

Sheffield, December 1982

TOPIC 10: EDUCATION AND TRAINING IN STATISTICS IN DEVELOPING COUNTRIES

ORGANISER: DR M N MURTHY

STATISTICAL EDUCATION AND TRAINING IN LATIN AMERICA: A GENERAL APPRAISAL

Enrique Cansado, CIENES, Santiago

INTRODUCTION

The problems that the teaching of Statistics generates are diverse and complex. They emerge either from the "formal" teaching (primary, secondary, and university), which is offered under the responsibility of the "academic statisticians", or from the training of governmental statisticians, who have responsibility for operation and supervision of technical and professional activities.

Both problems received rather early attention in Latin America. Because of its historical relevance we must point out the settlement of a career in Statistics in the University of Litoral (now University of Rosario) in 1948, in the city of Rosario, in Argentina. This pioneer effort was responsible for the training of a numerous set of professionals now active either in Argentina or in several Latin American countries. Another important early training program was developed by the Instituto Brasileiro de Estadística (IBGE), in Brazil. Both projects actively continue their development in our day. Some efforts have been accomplished in recent years to evaluate the development and evolution of the training in Statistics in Latin America. Among the most important ones, according to our judgement, must be mentioned "The Technical Meetings on the Evolution of the Statistical Profession in the Western Hemisphere", that took place in Santo Domingo, Dominican Republic (November, 1977) sponsored by the Inter American Statistical Institute (IASI), the "Latin American Seminar on the Teaching of Statistics" (SELENES), held in Santiago, Chile (November, 1977), organized by the Inter American Statistical Training Center (CIENES) and "Working Group on the Statistical Profession in the Western Hemisphere" gathered also at CIENES headquarters (December, 1980), under the sponsorship of IASI.

In the Session of the International Statistical Institute (ISI), held last year in Buenos Aires, Argentina, a "Round Table on the Teaching of Statistics in Latin America" was conducted and its proceedings will be published in a short period from now. At this event most of the relevant statisticians of the Region participated and they presented important contributions on the subject.

PRESENT SITUATION

The teaching of Statistics in Latin America presents today a situation highly superior to the one prevalent twenty or thirty years ago but the perspective of improvement is not as clear as it was then.

The countries of the Region are strongly diverse, as it is well known, in what concerns their sizes and general degree of development. The teaching of Statistics presents, as it is natural, different achievements and difficulties in these countries either in what concerns the academic area or in that related to the governmental one.

During a long period of time, and as a consequence of the Inter American programmes conducted with efficiency and continuity, by the Organization of American States (OAS) and the Inter American Statistical Training Center (CIENES), there was available for Latin America a set of training activities that complemented the efforts performed by the national institutions - either academic or governmental - related with teaching, research, extension, and application of statistical methodologies. A relevant part of this work was performed through the Inter American Center for Training in Economic and Financial Statistics (CIEF) from 1953 to 1961, and mainly by the Inter American Statistical Training Center (CIENES), that started in 1962 and is still active today, both located in Santiago, Chile. More than five thousand Latin American Statisticians have participated in the courses developed by these Centers in its headquarters or offered by their professors in the different countries.

This important and large amount of work has unfortunately been limited by the serious financial difficulties that OAS has suffered for some years.

As a consequence of this situation the countries of the Region, as far as the teaching of Statistics is concerned, have a better situation than the one they had in the past, but they will depend more and more on their own resources and initiative to develop the different programs in statistical training required in all of them.

In what concerns the formal education at primary and secondary level, no complete information is available and the little existent leads to the generalized impression that in the Region only a few incipient efforts have been performed as isolated experiments.

According to the available information it can be mentioned that thirteen of the twenty-eight state members of OAS have one or more university careers in Statistics; namely, Argentina, Bolivia, Brazil, Colombia, Costa Rica, Chile, Ecuador, México, Panamá, Paraguay, Perú, Dominican Republic, and Venezuela. In seven of these countries there exists, at the present time, one or more academic university programs at a graduate level (M.Sc. or Diploma) with strong differences in orientation, degree of consolidation, and number of participants. These seven countries are Argentina, Brazil, Colombia, Chile, México, Perú, and Venezuela. Only Argentina, Brazil, and México have programs at doctorate level founded and carried on by local universities.

All the countries recognize the importance of statistics teaching and the need for professionals in this field but they face serious difficulties in establishing such programs of education, perhaps the most important one being the lack of professionals with an adequate formation to offer teaching at a good level. This fact largely conditions the quality of several of the already existing programs. The few graduate programs existent in the Region make them unable to fulfill the present needs.

A characteristic of university teaching in this part of the world, and not extremely different from the one in the rest of the countries, is that in the formal teaching a heavy weight is given to statistical theory and a very light one to applications and actual data analysis; this fact, in many cases, does not enable the participants of these careers to work in offices of the national statistical system unless they make some complementary training.

Research in Statistics, either theoretical or applied, has not reached a significant level in Latin American universities. Research in statistical theory is conducted only in a few isolated places. On the other hand the interest for applied research is becoming more widely recognized. In spite of it, applied research faces at present many problems and the quality of what is being done varies very much among the different countries as well as among universities within the countries.

The countries of the Region are not different from the rest of the world in the fact that the university systems or other institutions of academic training have not provided or do not provide, at present, adequate training that would allow their participants to work in the production of basic governmental statistics. This reason has had as a consequence that several National Statistical offices - generally the heads of the National Statistical Services - have created their own Training Units in order to improve, at a basic and operational level, the present formation of their staff. In very few cases a good cooperation between academic and governmental statisticians exists; the gap between these two groups is far wider than the one existing in more developed countries.

As a very positive effort to solve this problem we should mention the recent creation in México, by a law dated December 1980, of the Instituto Nacional de Investigación y Capacitación en Estadística y Geografía (National Institute for Research and Training in Statistics and Geography), "with the goal of conducting research in order to improve the procedures, methods, and techniques in the field of statistics and geography as well as to offer training to the personnel in general and technical staff to satisfy the needs that an adequate integration and development of the national statistical systems requires".

In 1962 there was installed in Chile the Inter American Statistical Training Center, a dependent of the OAS, which has been offering training to professionals of the Region (either academic or governmental). Its contribution has been a significant one to this part of the world. In the academic area, the task has been reached through the Course in Mathematical Statistics, a graduate program at M.Sc. level, which has been completed by more than three hundred participants since it started. It was, for many years, the unique program of the Region offering a M.Sc. degree and still now is the course that produces the largest number of graduates at M.Sc. level per year in Latin America.

Several Latin American participants have completed doctoral programmes in American or European Universities. It is interesting to point out that almost half of them had previously obtained an M.Sc. degree at CIENES. The other two courses that CIENES regularly offered in its headquarters were dedicated to governmental statisticians. One was addressed to "producers" of statistics and the other one to "users" or "analysts" of statistics. More than one thousand four hundred Latin American governmental statisticians have profited from these courses. This significant amount of training activities has not a counterpart in the countries of the region.

CIENES has also been in charge of the Multinational Project of Educational Statistics since it was established in 1971. Every year courses on this field have been offered, and about two hundred students have participated in them.

More than one thousand five hundred students have participated in other specialized courses offered by CIENES.

The fields in which these courses were offered cover a very wide range of subjects: Agricultural Statistics, Trade Statistics, Statistics of Tourism, Demographic Statistics, etc.

CIENES has also offered courses in different countries of the region. Sometimes it has also participated in already established training programs sending one or several of its professors. About three thousand Latin American statisticians have participated in these courses, some of them academic ones but the major number in the governmental area.

CIENES continues offering at present courses for all the Latin American countries in its headquarters. Unfortunately the number of scholarships that OAS grants is extremely reduced. In this last year CIENES has dedicated a serious effort to support its Program of Assistance to the Institutionalization of the Teaching of Statistics (PRAIENES), that has as its goal, to offer help to the national institutions in the different countries of the Region, which have regular problems in the teaching of statistics, in order to strengthen such programs or to create new ones.

SOME COMMENTS ON THE PRESENT SITUATION

There is little correlation between the degree in which the importance of Statistics is accepted in an early stage of a process of education and how much has been implemented in this direction at primary and secondary level. So it seems quite commendable to make efforts to strengthen the teaching in this field, when it exists, or to establish it when it is not already offered. In order to accomplish this, well qualified professors must be available and, even before that, well elaborated programmes with a clear enumeration of well selected topics must be prepared. The intuitive aspects of Statistics and its possibilities of application should be clearly pointed out in order to emphasize the difference between statistics and classic Mathematics Problems. Problems like data production, inference, and processes of taking decisions should be presented. Given the complexity and wideness of the problems that the teaching of statistics at primary and secondary levels raises a close collaboration between statisticians and experts in education will be necessary as well as the participation of all national and international institutions involved in the problem. At the ISI Session held last December in Argentina, a round table on the problem of the teaching of statistics in Latin America was held.

It was pointed out there that the difficulties involved in keeping the university programs at an adequate operational level in almost all the countries was caused by economic reasons. The need of an evaluation of the labor market for the graduates of careers in Statistics was also discussed.

It is strongly expected that this Conference will be able to produce some recommendations on the problem.

In December 1980 a Working Group on the situation of the Statistical Profession in the Continent met at CIENES. As a consequence of this meeting very valuable recommendations on the teaching of statistics at University level were made.

These recommendations follow:

1. It is absolutely necessary to continue and to expand educational and training programs in statistics at the graduate level. This would have a twofold effect: It would increase the amount and the quality of research being carried out, and it would have a multiplier effect as well through the use of these professional statisticians as teachers. It should be noted that the quality of instructions improves if the teachers do research work.
2. To diminish the problem of financing research, advisory services, and manpower training for these activities, the groups of researchers should prepare projects in areas of primary interest to government agencies or industry. This point is particularly relevant for researchers working in the academic field.

3. In the field of international technical assistance, and especially where advisory and research services are involved, agencies sponsoring such services must increase their financial support. As a counterpart for this, the organizations that carry out this assistance must propose projects that are sufficiently feasible and significant to attract the interest and support of the governments of the region.

4. Given the great scarcity of publications in the Latin American area, communications among the various groups of researchers is difficult. Efforts should be made to have a regular periodic publication that would essentially benefit research in the region.

5. The more frequent and regular scheduling of national and international meetings would also be desirable. Meetings should be designed to favor the exchange of data on experience gained by researchers, and an interchange of information between researchers and professional users of results derived from research.

6. Research and advisory services should be institutionalized and be mandatory in both governmental and academic organizations. These activities should also be upgraded to be significantly rewarding to the professional at the economic and social level.

7. An active exchange among already existing research groups should be encouraged through periodic visits, exchange of professors, and other similar activities.

A good graduate program in Statistics at M.Sc. level would, as a first step, provide professors with an adequate amount of knowledge to be able to teach in university careers in Statistics. This M.Sc. program can be followed inside the same country, in another country of the Region (in most cases with the advantage of receiving the teaching in the own language of the participant) or in a university outside the Region.

There are very few doctoral programs in Statistics in the Region. Some of them have a serious prestige, but it could be said that at that level almost all the training is followed by Latin American candidates at American and European Universities. In several Inter-American meetings the first ideas of a project on a degree at doctoral level jointly sponsored and granted by some universities of Argentina, Brazil, and México and coordinated by CIENES has been discussed. However desirable the project may be, it seems not to be an easy one and its implementation cannot be thought about in the near future.

If the problems that arise in the process of improving the formal teaching of Statistics are varied, diverse, and difficult, the ones that face the improvement in the capacity of governmental statisticians in Latin America are even more complex. Given the deep gap that separates (in the Region, even more than in other parts of the world) academic and governmental statisticians, the training of these last ones is not offered by the universities but through international centres. Examples of these last ones are CIENES and, with the same location, the Latin American Center for Demography (CELADE) depending on the United Nations. The National Statistical Offices of the Region are more and more establishing programs of Training to satisfy immediate objectives, but only at a basic or operative level.

SOME STATISTICAL FACTS

On the occasion of the 43rd Session of the International Statistical Institute, Dr. Raúl Mentz and Mrs. Nora M. Jarma de Cortés presented an invited paper [5], which will be the main source of some specific information about the situation of statistics as a profession in the Region.

The paper of Dr. Mentz and Mrs. Jarma de Cortés defines a quite enlightening line of research and CIENES will dedicate some efforts to it in the future. Dr. Mentz and Mrs. Jarma de Cortés belong to the National University of Tucumán, in Argentina. The information derives from a sequence of surveys implemented since 1979. The original frame of 27 centers was complemented with the list of institutions invited to the Latin American Seminar on the Teaching of Statistics (SELENES) gathered at CIENES in 1978, leaving the final number of consulted centers at 106. The answer came only from 29 Latin American Centers, which belong to Argentina (6), Brazil (11), Colombia (3), Chile (4), Guatemala (1), México (3), and Venezuela (1).

Table I of the above mentioned paper is reproduced on the next page.

The survey conducted by Professors Mentz and Jarma de Cortés is a serious effort but the non-response to the questionnaire is reflected in several absences. To give a few examples: There are several groups in Chile which are not recorded; they are located in Santiago, Valparaiso, Iquique and Antofagasta. Two strong Colombian groups are also absent (National University of Bogota and del Valle University in Cali). In the Caracas area are also several groups. This comment must not be considered a non-constructive criticism, it is just made in order to express that it is CIENES intention to present at the ISI 44th Session, which will be held in Madrid in 1983, an inventory of the Latin American statistical situation as complete as possible. The short period of time that elapsed between the ISI 43rd Session and the moment at which this paper is written made it impossible to conduct the required survey.

References

[1] Documentos presentados en la Reunión Técnica sobre la Evolución de la Profesión Estadística en el Hemisferio Occidental (Noviembre, 1977), Instituto Interamericano de Estadística (IASI), Washington, D.C., U.S.A.

[2] Seminario Latinoamericano sobre Enseñanza de Estadística (Noviembre, 1978), Centro Interamericano de Enseñanza de Estadística (CIENES), Santiago, Chile.

[3] Adiestramiento Estadístico en la Región Americana (Agosto, 1979), Centro Interamericano de Enseñanza de Estadística (CIENES), Santiago, Chile.

[4] Informe de la Reunión del Grupo de Trabajo para examinar la Situación de la Profesión Estadística en el Hemisferio Occidental (Diciembre, 1980), Instituto Interamericano de Estadística (IASI), Washington, D.C., U.S.A.

[5] Mentz, Raúl Pedro and Iarna, Nora M. de Cortés (1981) Encuesta a Centros Estadísticos Latinoamericanos (Survey on Latin American Statistical Centers). Invited paper 43rd Session of International Statistical Institute, Buenos Aires, Argentina.

[illegible]

BEST COPY AVAILABLE

STATISTICAL EDUCATION AND TRAINING IN THE DEVELOPING COUNTRIES IN AFRICA

J B Adichie and Biyi Afonja
University of Nigeria, Nsukka and University of Ibadan

INTRODUCTION

It is now widely recognised, if still grudgingly by some academicians and persons in older professions, that statistics is both an academic discipline and a profession. In Africa as indeed in every other continent of the world, the teaching of statistics may be conveniently put into two broad categories: the teaching of statistics as an academic discipline in schools and colleges and the training in statistics as a profession. It would therefore be convenient to treat the topic of this paper in two parts, namely: statistical education corresponding to teaching statistics as an academic discipline in institutions and statistical training for a profession.

STATISTICAL EDUCATION IN AFRICA

Until recently, statistics was not considered pure or important enough to be admitted as an academic discipline in its own right. This phenomenon was not peculiar to developing countries, indeed it was even more pronounced in the developed countries of the world. Because the school and college curricula of most African countries were faithfully modelled after the colonial patterns, the type of statistics taught in schools then was made a part of the mathematics syllabus. In schools, the course consisted mainly of elementary techniques in the calculation of means and variances, binomial coefficients and Poisson probabilities. Feeble attempts made to introduce into the school syllabus practical and useful, yet elementary statistical topics like sampling procedures, methods of censuses and surveys, statistical reasoning, were quickly killed by purist mathematicians who would not compromise the elegance and beauty of mathematics by admitting such topics that are deemed to lack mathematical rigour. In colleges and universities, where it was taught at all, statistics was treated as part of Applied Mathematics. This was the state of affairs in most institutions in Africa in the sixties.

Teaching Statistics in Schools

In the early seventies, however, pressure was brought to bear on the schools to cultivate statistics as a distinct school subject. This was the case mainly with the schools in Anglophone countries of West Africa. Two principal factors were responsible for the pressure; the first was the crying need to have a recruitment base for the statistical assistants who would fill the many vacant positions in the National Statistics Offices of these countries. It must be remembered that about this time, many African countries had gained independence from their erstwhile colonial masters and were therefore very mindful of consolidating their economy in this post-independence period. It was therefore important to strengthen the national statistical offices, by recruiting not only the professional and middle level statistical manpower but also the junior statistical personnel. A secondary school graduate with the General Certificate of Education (GCE) in Statistics and Mathematics was a most fitting candidate for these junior positions of statistical assistants.

The second factor which is related to the first was that the National and Regional Examining Bodies, e.g. the West African Examinations Council (WAEC), agreed to examine statistics in the school and the GCE 'O' level examinations. Appropriate syllabuses had to be worked out, discussed and approved. The WAEC syllabus for the GCE 'O' level and West African School Certificate (WASC) statistics contains the following topics among others:

> collection of data; sources of data, planning of censuses and surveys, random sampling, design of questionnaires, tabulation and presentation of data; diagrams and charts, measures of location and dispersion for a set of numbers and for grouped data. Calculations of simple probabilities simple uses of the binomial and normal distributions. Weighted averages; index numbers. Simple regression and correlation.

This syllabus is run by all the English speaking West African Schools that offer statistics in the GCE 'O' level examination conducted by WAEC. At the moment these are schools in Nigeria, Ghana, Sierra Leone, Liberia and the Gambia. A similar syllabus also exists for the East African Examining Bodies.

Comparative figures showing the number of Nigerian candidates offering statistics and those offering other subjects like Mathematics, Economics and English Language in the GCE 'O' level conducted by the West African Examinations Council are given in the following table.

Nigerian Candidates Entry for GCE 'O' Level Examination

Subject	Year 1978	1979	1980	1982
Statistics	1,335	1,486	1,665	1,864
Mathematics	119,486	142,909	176,221	316,979
Economics	104,096	128,704	159,841	291,260
English Language	123,583	148,317	174,259	316,979

Source: West African Examinations Council

This state of affairs is indicative of the position in the entire continent South of the Sahara. Our experience even shows that Nigeria's embrace with school statistics is much warmer than that of other countries.

Facilities for Statistical Teaching in Schools: One of the visible constraints in the development of school statistics in Africa is the lack of qualified teachers. It is discovered that even in the schools that offer statistics in the GCE 'O' level examinations, only a few teachers can really teach the subject. One typical example is found in a Nigerian School where 240 pupils (i.e. 6 classes of 40 each) are taught the subject for two years, by only one statistics graduate teacher! Obviously, most of the time, the students fend for themselves reading whatever is available in the nature of books. And these books are either not there or very scanty indeed.

Lack of suitable text books for school statistics is one of the commonest problems threatening the teaching of school statistics. But this problem has a special dimension for Africa, in the sense that most of the statistics text books used in institutions in Africa are written and published outside Africa, mostly by non-Africans. When there is inadequacy of books in the countries of the authors, its effect on the countries on the receiving end can be imagined. Right now only one or two books specifically written by Africans are near enough to what is needed.

Prospects of School Statistics: It should be noted that with all the arguments in favour of introducing statistics as a school subject, some reservation is still expressed in some quarters. The principal argument is that strengthening the mathematics base of students in schools should be preferred to adding statistics as a subject which itself would need mathematics for further development. The result is that even with the introduction of statistics as a separate examination subject some statistics continues to be taught as part of school mathematics. It is, however, hoped that by the time recommendations of Vic Barnett's ISI (International Statistical Institute) sponsored Task Force on the teaching of statistics at school levels (TOTSAS) are implemented, many more countries in Africa will have firmly established the teaching of statistics in their schools.

Teaching Statistics in Higher Institutions in Africa

The gaining of political independence and the subsequent expansion of their economies have made it necessary for many African countries to seek to strengthen their government statistical services with respect to professional and middle level manpower. A few higher institutions were encouraged to start undergraduate and in some cases, sub-degree courses in statistics either as a separate subject or as a part of mathematics or economics.

The first International "Survey of Statistics Teaching in developing Countries" carried out by UNESCO in collaboration with the ISI in 1971/72 reported that only twelve University-level institutions in the following ten African countries had programmes leading to a qualification of up to a first degree level in statistics as at 1972:

> Ethiopia (Cairo and Alexandria), Ethiopia (Addis Ababa), Ghana (Legon), Ivory Coast (Ecole de Statistique d'Abidjan), Morocco (Institut Nationale de Statistique et d'Economic Appliquee, Rabat), Nigeria (Nsukka and Ibadan), Sudan (Khartoum), Tanzania (Dar es Salam), Uganda (Makerere), Zaire (Lumumbashi).

Among all these African institutions, there were in the 1971/72 academic session, about 230 students in their final year of the first degree programme or equivalent, and over 300 postgraduate students mainly in Cairo and Khartoum pursuing post-graduate diploma and M.Sc. programmes, with a total of 135 full time equivalent teachers.

Apart from the above mentioned institutions that offer degree level courses in statistics, there are other institutions that offer programmes, notably in mathematics and social sciences that contain substantial and often compulsory statistical components. Many institutions also offer service courses in statistics to students of social sciences, engineering, medicine and agriculture. The survey report recorded 33 such institutions in Africa as at the 1971/72 session.

As at 1981, the number of higher institutions in Africa offering degree-level courses in statistics has tremendously increased. For example, in Nigeria alone, the number of universities offering degree programmes in statistics has risen from two in 1972 to ten in 1981. These are at Benin, Calabar, Ibadan, Ife, Ilorin, Lagos, Maiduguri, Nsukka, Port Harcourt and Zaria.

Besides the increase in the number of university level institutions in each country, there is also a noticeable increase in the number of countries that now have institutions offering degree level courses in statistics. Universities in Botswana, Lesotho and Swaziland, and special institutions in the Cameroun, Algeria and Rwanda are recent examples.

Programme in Universities: The programmes in these universities particularly in Nigeria are offered in various forms. For example, there are departments of statistics in Nsukka, Ibadan and Ilorin, while in the other universities the statistics programme is offered in the department of mathematics.

In addition, Ife and Benin have some statistics courses located in demography/ social statistics and economics departments respectively.

By and large, the B.Sc. statistics programmes of most of the universities, particularly those in the Anglophone countries, appear similar to the standard B.Sc. statistics programme of either the British or the American/Continental University. There is enough of mathematics - calculus and linear algebra - in the syllabus to help the students understand the basic probability theory required for a thorough grasp of statistical methods.

A typical syllabus for a B.Sc. (Statistics) would have courses built around the following topics:

> Probability and distribution theory, statistical inference, sample surveys, design of experiments, demography, actuarial statistics, statistical methods; non-parametric methods, multivariate analysis, econometrics, and sequential analysis, economic statistics, time series analysis, sources of official statistics, operations research and practical project.

The balance between theory and practice (Applied Statistics, an elusive term) varies among the Universities depending on the needs of the country, the orientation of the lecturers and the facilities available.

In some of the institutions, it appears that efforts have been made to modify the 'imported' programme to suit local conditions. Even then, as in the developed countries the standard B.Sc. statistics programme continues to be criticised as being inadequate for preparing graduates for careers in government. Recent discussion at the 9th Conference of Commonwealth Statisticians held in Nairobi in December, 1980 seems to suggest that this criticism is not limited to developing countries only. The B.Stat. programme run parallel with the B.Sc. at Makerere University is obviously another attempt to meet these criticisms.

A few universities: 3 in Nigeria, 2 in Egypt, one in Sudan, one in Ghana and one in Uganda now offer postgraduate programmes in statistics leading to Master and Doctoral degrees.

Programme in Colleges of Technology and Education: Apart from the universities per se, there are also particularly in Nigeria, many polytechnics, colleges of technology, and colleges of education that include statistics in their programmes. Prominent among this group of higher institutions in Nigeria are: The Polytechnics in Ibadan, and Kaduna, the Institute of Management and Technology in Enugu; the Colleges of Education at Owerri, Awka and Uyo.

The standard of statistics education offered in these colleges is indeed post-secondary if slightly sub-degree. The polytechnics and colleges of technology award Ordinary National Diploma (OND) in Statistics, while a few go up to the Higher National Diploma (HND) after four years of study. The qualifications are very close to the B.Sc. level, but are usually practical in nature. Particular mention should be made of the programme of Colleges of Education. The graduates of these Colleges of Education are awarded a Certificate in Education with emphasis on the teaching subject studied. Those of them that study statistics (with mathematics usually) get enough of statistics to be able to teach the subject in secondary schools. Some colleges of education now offer additional university type degrees.

Francophone Centres: Institutions offering statistics in the French speaking African countries would for the purpose of this paper be classified as Training Institutions different from the standard University. We shall consider them in more detail in the next section.

TRAINING STATISTICAL STAFF

So far, the discussion has dealt mainly with the teaching of statistics as an academic discipline. Except at the lowest (school) level, teaching statistics for general education may also be seen in the light of preparing staff for professional (statistical) duties.

To meet the needs of governments, special training institutions have been set up particularly in the Francophone countries.

Before examining the training of this form, it is relevant to mention the type and nature of work expected to be carried out by statisticians and their subordinate staff. Broadly, the work falls into three categories:

(a) data collection, processing, analysis and dissemination carried out in government and quasi-government offices.

(b) research of the type carried out in formal research centres or similar organisations.

(c) academic, mainly teaching with possibly some research.

Corresponding to these three categories are three types of statisticians styled by Healy (1973) as intelligence statistician (a pseudonym for official statistician), experimental statistician and statistical mathematician respectively.

It is the training of official and experimental statisticians that engages the attention of many African countries. It is relevant to mention at this stage the intensified efforts of the United Nations Development Programme (UNDP) through the Economic Commission for Africa (ECA). At the instance of the conference of African Government statisticians (now succeeded by the Joint Conference of African planners, statisticians and demographers) a programme called the Statistical Training Programme for Africa (STPA) was started at the ECA in 1978. The main objective of the programme is to ensure adequate supply of trained statistical personnel through the coordination of the activities of the institutions being helped by the UNDP.

The Centres in Ibadan, Legon, Lesotho, Botswana and Makerere have already been mentioned because they are run within university settings. The other centres are in Yaounde, Abidjan, Algiers, Rabat, Kigali, Dar es Salam and Paris. Details about these centres can be found in the ECA document Directory of Centres participating in the STPA.

Junior Cadre

Many National Statistical offices now offer internal training for enumerators, statistical assistants/clerks, who usually have GCE O-level or equivalent in mathematics as their basic entry requirement. Right from this level, economics and mathematics are offered along with statistics. The aim here is to train the assistants in the practical (not theoretical) aspects of their job. It is to be mentioned that any of the assistants with GCE O-level in statistics need not undergo this training.

Intermediate Level

In view of the traditional approach of the few existing universities in Africa and realising the need to produce enough statistical manpower, special middle-level training programmes were instituted in Accra, Ibadan, Dar es Salam, Rabat and Yaounde in the early and mid sixties. The programmes in Accra, Dar es Salam, Rabat and Yaounde were handled through independent training centres whilst the Ibadan programme was run as a special programme within the Faculty of Social Science of the University of Ibadan. Though rudiments of such courses as experimental design featured in some of the curricula, the intention was clearly on the training of what can be called middle-level economic (and to a lesser extent, social) statisticians, the prospective official statisticians.

All the centres had the support of the UNDP through the ECA and the UNESCO. The centre at Ivory Coast now also offers training at this level as an independent centre. In the English speaking countries the centres in Lesotho, Botswana and Nigeria offer certificate and diploma programmes (at intermediate levels) within the context of the programme of the University to which each is attached.

Centres in the French speaking countries offer programmes leading to Agent Technique and Adjoints Technique Diplomae. A centre at Lubango, Angola offers training for the Portuguese speaking countries.

There are also other training centres, not under the UNDP, which offer training at this level. An example is Ecole Nationale d'Economie Applique, Dakar, Senegal.

Curriculum for Intermediate Training: As earlier indicated, the curriculum of the centres was designed to train middle level official statisticians. Though the bulk of the graduates of these centres will by and large serve as such, some flexibility in the curriculum is now being attempted, because it is realised that the products of the centres may not all work in government statistics offices and that the more able ones among them may need to pursue higher studies.

For the Francophone centres a typical Agent Technique syllabus has the following as its core subject groupings: General studies comprising mathematics, numerical computation, French, English, economics and geography; basic statistics, applied statistics, computer science and practical project.

The Adjoint Technique syllabus has additional topics like statistical organisation and preparation; elements of civil and commercial law, development and planning statistical analysis, survey theory and practice.

The Anglophone syllabus falls broadly under Economics, including economic geography; mathematics, statistics theory and methods, applied statistics, practical project.

While statistics, mathematics and computer science are seen as core subjects in these programmes, the exact balance among them is still open.

It is, however, understood that the ECA is working towards the production of a guide syllabus flexible enough for adaptation at most African institutions.

Professional Level

All the five Francophone centres in Africa (Abidjan, Algiers, Kigali, Rabat and Yaounde) offer various degree equivalents in statistics, Ingenieur des travaux statistique (ITS), economic statistics, Ingenieur statisticien economistes (ISE), applied statistics, Ingenieur d'application de la statistique (IAS), and in one case information science, Ingenieur analystes information (IAI). The Centre Europeen de formation des Statisticiens, economistes de pays en Developpement (CESD) Paris was for a long time solely responsible for training Francophone Africans at the ISE level. It continues to do so alongside the growing centres in Africa.

In the English speaking countries, no special independent training centres exist as such, other than those already mentioned in the section 'Teaching Statistics in Higher Institutions in Africa'.

Curriculum for Professional Training: A typical ITS syllabus would have courses similar to those contained, in the B.Sc. (Statistics) syllabus given earlier, but with more pronounced courses in economics and planning e.g. economic theory, world economic systems, national accounts projects selection and evaluation theory. A lot of weight is attached to the practical work that normally follows classroom lectures. Lectures and practical work in the major subject areas are spread over the period of training according to some relative weighting, e.g.

mathematics 15%, statistics and demography 31%,
economics 28%, computer science 10%, languages 8%,
miscellaneous and case studies 8%.

Special Training: To produce further training for serving statisticians, postgraduate courses, seminars and workshops of varying duration are now being run in some institutions and centres.

For example, in Ibadan a one-year postgraduate diploma is being run primarily for serving statisticians who are non-statistics graduates.

PROBLEMS AND PROSPECTS

We conclude this review by highlighting some problems inherent in statistical education in Africa. An insight into the future is then undertaken.

Because the educational systems of the two main foreign language groups (Anglophone and Francophone) are different no excessive generalisations can be attempted. Nonetheless we examine the problems under the headings: curriculum and standards, staffing, facilities, organisational set up and manpower needs.

Curriculum and Standards

In the context of the work of official statisticians, the syllabuses of the Francophone centres at both middle and high levels seem to be broadly based. While statistics, mathematics and economics remain common topics in all the programmes, the type of degree awarded appears dictated by the emphasis. Thus we have ITS, and ISE being awarded within the same institution. This practice obviates the problem of finding a suitable name for a degree in which some depth in statistics has to be sacrificed for other subjects. However, with such other subjects as business, law, accounting being an integral part of the syllabuses, a kind of general degree would be implied. Perhaps this is why statistical education in the Francophone countries is not normally part of a university setting.

In the Anglophone countries, there seems to be some confusion between the middle level and higher (degree) level training. From the foregoing account, one can notice an attempt to 'professionalise' the training at the middle level. Though many of the universities mentioned attempt to practicalise their degree programmes, yet this practicalisation is not to the same extent and level as in the Francophone institutions.

If such professionalisation were attempted within the existing university set up, many problems of logistics would arise. For example without a well established course system, a department of statistics trying to offer courses in national accounts may run into difficulties. With a course system, however, such topics can be offered in appropriate departments.

The Makerere experiment of running a parallel B.Stat. programme is yet to be evaluated for its advantages over a well designed B.Sc. In summary the basic question that still remains is that of determining how much of applied statistics topics (economic statistics, educational statistics, agricultural statistics) and miscellaneous topics (national accounts) should be taught in universities and training centres. On the other hand, it is relevant also to ask how far African statistics departments and training centres should go in teaching topics like probability theory, theory of sample survey and design of experiment as distinct from their applications. We do not pretend to have the answers. It seems clear, however, that more emphasis is likely to be placed on additional applied options even at the expense of the theoretical base. Suffice it to say that with continued discussion and practice, the right mix will appear at the appropriate time.

Staffing

With a normal statistics unit (department) offering standard statistics courses, lack of staff is a near universal problem. From our experience of universities in and outside Nigeria, we doubt if there is a single university in Africa which has enough qualified African staff to man its statistics department adequately. The problem becomes worse if departments are to teach some of the various applied options already mentioned. We have noticed in one or two instances when, even among the available staff, nobody was willing to teach some particular applied courses for reason of lack of either text books or interest.

We realise that a practice for combating these problems is to hire short term specialist staff from outside the institution itself. If well organised this may well be a way out. But it can be no substitute for tenured long term experienced staff.

It is pertinent to reiterate the help that UNDP has been giving through UNESCO and the ECA to many of our institutions. Many other multilateral and bilateral agencies have been assisting too. It appears that such help will be needed for some time to come.

Facilities

Lack of adequate text books and teaching materials has already been mentioned. The problem is all the more acute in areas where texts are needed to suit local conditions, for example official statistics (or at least topics in its methodology) and national accounts.

Other facilities that are usually not at hand are mainly computing (calculating) machines. The effect of this on what is taught is clearly obvious.

Organisational Set Up

We have noticed three or four organisational hybrids for teaching statistics through

(a) conventional statistics departments in universities or polytechnics;
(b) an existing university department like mathematics or economics;
(c) institute of statistics within university set up and;
(d) independent training institutions awarding their own diplomas, degrees, etc.

In the discussion under curriculum and staffing, mention has been made of the associated problems under each set up. It may be added that in some universities both situations (a) and (b) exist for one reason or the other. Whatever operates in any given institution in a country will certainly depend on many factors that sometimes have nothing to do with statistics.

Manpower

Perhaps a most disturbing fact is the slow output rate of statisticians from all the institutions. Measured against expected demand by 1987, which according to an ECA survey, is put at about 5,000 at middle level and 7,000 at professional level, the present output rate will hardly ever catch up with these demand figures. Even with the low output, not many of the statistics graduates from universities take up employment in public services statistical agencies. It may well be that the professionalised training centres are to be preferred to the university, as the graduates of the latter seem to have greater opportunities and wider avenues to choose from.

References

[1] Healy, M.J.R. (1973) The varieties of statistician. Journal of the Royal Statistical Society, Series A, 136, 71-74.

[2] 9th Conference of Commonwealth Statisticians. Summary of Proceedings, (1980) Nairobi.

[3] Directory of Centres participating in the statistical training programme of Africa. (Preliminary edition) ECA. (1981)

[4] Survey of statistics teaching in developing countries. UNESCO (1971).

STATISTICAL EDUCATION AND TRAINING IN DEVELOPING COUNTRIES OF ASIA AND THE PACIFIC

M N Murthy, Statistical Institute for Asia and the Pacific, Tokyo

INTRODUCTION

There have been considerable changes in the developing countries of Asia and the Pacific in the last three or four decades in the context of economic and social development in the post-war era. Statistics is increasingly being used in studying and finding solutions to problems of change. In this connexion, statistics has become an indispensable new technology for use in development efforts and assessment of gains due to such efforts. This has led to the governments of developing countries establishing and expanding statistical offices to meet the growing demands on statistics. Recently the semi-government and private sectors have also expanded reasonably well in these countries providing further opportunities for use of statistics. All these developments have invariably resulted in a growing need for trained and experienced statistical manpower to meet the challenges of change and development.

For a balanced development of statistical personnel it is important to develop both statistical education and statistical training facilities. It would have been desirable to have developed the statistical offices in the developing countries after establishing statistical education facilities to provide the necessary statistical methodology. However, the force of changes in economic and social development had necessitated the establishment and expansion of statistical offices even before statistical education and training facilities could be created and stabilised. The last two or three decades have seen considerable development of statistical education and statistical training facilities in the developing countries of Asia and the Pacific including establishment of regional statistical training centres such as the International Statistical Education Centre (ISEC) at Calcutta and the Statistical Institute for Asia and the Pacific (SIAP) at Tokyo.

Since its inception, SIAP has been interested in the training requirements of the developing countries of the region and has carried out surveys in the past for this purpose. Since statistical education and training are both required to produce effective statisticians with sufficient background knowledge of theory and practice, SIAP has recently conducted an enquiry on statistical education and training facilities in the developing countries of Asia and the Pacific. The results of this survey are reflected in this paper and the details are covered in the section on 'Reports from Countries'.

The main aim of this paper is to assess the present position regarding the extent of the availability of statistical education and training facilities in the developing countries of the ESCAP region and to discuss some of the main problems being faced by these countries in developing the statistical manpower needed in the context of the development efforts. Early developments in the fields of statistical education and statistical training are briefly covered in the section 'Early Developments' and university education in statistics, regional and national training facilities are taken up in the sections 'University Education in Statistics', 'Regional Training Centres' and 'National Training Centres'; teaching of statistics in schools is briefly discussed in the section 'Concluding Remarks'. No attempt is made in this paper to review or evaluate the design and contents of the courses offered by the universities, institutes and training centres in the region.

Statistical education, training and research have been discussed at length by several authors and by committees set up by national and international organizations. Montenegro (1973) reviewed the efforts at promoting statistical education and training and referred to reports of various committees and conferences in this connexion. Bjerve (1976) in his presidential address at the Warsaw session of the International Statistical Institute brought into focus the gap that had developed between theoretical and practical statisticians and the problems relating to the inceasing specialization in various subject-matter areas in statistics. Since then, the SIAP, the American Statistical Association (ASA) and the ISI had set up panel discussion, conference and committee to discuss these problems, and they have come up with many useful recommendations (Murthy and Khan, 1977; ASA, 1978; Duncan and Durbin, 1980). ASA (1978) includes many useful papers on statistical education and training and particular reference may be made to the papers by Fellegi, Malinvaud and Murthy. United Nations (1979) has discussed human skills and training in the context of statistical work. Finker and Washabaugh (1977) have described many programmes available for training domestic and foreign government statisticians in the U.S.A.

The ISI and the UNESCO have been engaged in the promotion of statistical education, whereas the United Nations and its specialized agencies such as FAO, WHO and ILO, have been concerning themselves with the promotion of statistical training and standardization in data production. This division of work has evolved over the past 30 years without appreciable co-ordination of the activities, leading to some problems in relating statistical education to work requirements, and in turn giving rise to the situation referred to above. This is further discussed later in this paper.

EARLY DEVELOPMENTS

Though statistics has a long history in its practical aspect in the sense that data collection and compilation had been of interest to administrators from the very early times, statistical methodology is relatively of recent origin and owes its development and promotion to persons like Karl Pearson, Ronald Fisher, Jerzy Neyman and P.C. Mahalanobis. In the Asia and Pacific region, the promotion of statistics and the development of statistical education and training have taken place in most of the developing countries during the last three decades of the post-war period. India has had a longer association with statistics and its development, and ranks among the relatively more developed countries in the field of statistics.

In India, there is a record of a statistics committee being set up in 1862 to propose modern statistical forms for collection and compilation of statistics in the fields of trade, finance, education, agriculture etc. which later led to the publication of statistical abstracts in 1868 (CSO, 1970). With the establishment of the Indian Statistical Institute in 1931, statistics became accepted in India and the universities gradually started offering courses in statistics, initially as part of departments of mathematics and later as separate departments of statistics. This process of statistical development was strengthened by the founding of the Indian Agricultural Statistics Research Institute in 1947 and the Central Statistical Organization in 1951. The development of statistics in India owes much to the dynamic leadership and sustained efforts of P.C. Mahalanobis.

The Statistical Office of the United Nations and the Statistical Sections and Divisions of the Specialized Agencies such as FAO, ILO, WHO and UNESCO have done commendable work in getting statistics accepted in the developing countries and in promoting and effecting statistical development by organising international technical assistance programmes. Standardisation of concepts, definitions and tabulation plans in population, housing, agriculture and industrial censuses has been advanced by the United Nations with the co-operation of all the countries including the developing ones. The United Nations Statistical Commission and its Sub-Commission on Statistical Sampling established in 1947 have done much to promote and develop statistics in the developing countries of the world, (United Nations 1947-51). Even now the United Nations and its specialized agencies are active in this field and there is currently the National Household Survey Capability Programme (NHSCP) of the United Nations which aims at setting up continuing household surveys in the developing countries of the world.

In this region, the ESCAP has been promoting and developing statistics in the developing countries through the organisation of working groups, seminars and technical meetings on various aspects of data production and utilisation as well as through technical assistance to these countries. In some countries, expansion of statistical data collection and compilation grew faster than the development of statistical education facilities leading to the employment of personnel lacking in suitable background for actual statistical work. Further, the universities initially started teaching statistics as part of mathematics department courses and naturally emphasised the theoretical aspects of statistics with a mathematical orientation at the expense of statistical operations and applications. This led to the gaps between theory and practice and between theoreticians and practitioners mentioned in the Introduction. This imbalance in statistical education and work requirements, which developed in the early period and persists even now, has necessitated the creation of national and international training facilities to remedy the situation. Recently there has been considerable constructive discussion on the problems of statistical education and statistical training and the current situation is discussed further later in this paper.

The situation mentioned above had been realised by the United Nations (1949) and by the ISI (1951) in the early stages. For instance, the Statistical Commission in its report in 1949 mentioned that it recognized there was a shortage of adequately educated and trained statisticians in many countries and that the burden of statistical work in public and private institutions was large, would increase and would have to be met by the education and training of more and better statisticians. The ISI has taken many steps to strengthen statistical education, one such step being the setting up of the International Statistical Education Centre at Calcutta in 1950. Also, the ISI has been holding round-table conferences on teaching of statistics and meetings of Directors of International Statistical Education Centres in the recent years. However, the present International Conference on Teaching Statistics being held at Sheffield in August 1982 under the sponsorship of the ISI is the first of its kind, as there has been no such conference so far in the world.

UNIVERSITY EDUCATION IN STATISTICS

Even in the early years of development of statistical education in universities considerable attention has been given to this topic as is evidenced by the report of the Council of the Royal Statistical Society (1947) and the report of the Commission on Teaching of Statistics of the Institute of Mathematical Statistics (1948). More recently, the ISI (1969) organized a round-table discussion on university teaching of statistics in developing countries. UNESCO (1974) has surveyed the situation regarding teaching of statistics in developing countries. The problems of teaching statistics at university level also get reflected in many seminars and conferences including the ISI biennial sessions. The ISI Committee on Integration of Statistics had also considered the problems of statistical education in universities (Duncan and Durbin, 1980). A useful discussion of teaching statistics in universities has been given by Daniels (1975). The periodical American Statistician has a teachers' corner to discuss problems of teaching statistics and the periodical Teaching Statistics provides a very good forum for teachers to exchange ideas on course designs, course contents and methods of teaching. Also a number of teaching aids in the form of cassettes, slides, movies and video-tapes are becoming available, and their proper use will help increase the effectiveness of teaching statistics.

In the Asia and Pacific region, except India where university education in statistics developed much earlier, the development in this connexion has been of a recent origin in many developing countries and in some countries the university facilities are still non-existent or available only at a minimum level. Historically, statistics came under the Department of Mathematics in many universities and this naturally led to an emphasis on theoretical aspects of statistics with a mathematical orientation as mentioned earlier. Though statistics is basically an applied subject, it was and is being taught in many universities as a branch of mathematics. The graduates coming out of universities become the teachers of statistics and research workers, and this inbreeding ensures continuation of the stress on the theoretical aspect in teaching of statistics and research work in this field.

University education in statistics concentrates on the so-called mathematical statistics and even when applied subjects such as sample surveys and design and analysis of experiments are taught, the emphasis is on their mathematical orientation. However, after graduation most of the students generally find employment in statistical offices of government and semi-government agencies and in business and industry, where they have to design and carry out the statistical operations of a statistical investigation including data collection, compilation, analysis and interpretation, still not generally covered at all or covered in the universities with considerable abstraction.

The above situation in this region has been noted at a very early stage of development of statistical education in universities by Madhava (1951) when he said we had failed to take adequate care of practical work in statistics. The United Nations (1981, pp.42-43) states that in many countries the knowledge acquired in universities, both in subject-matter fields and in statistics, was often too general and abstract to be immediately well adapted to the needs of statistical agencies. It may be mentioned that it is desirable to take the general objectives of statistical education as developing a statistical sense with a view to understanding natural and physical phenomena through statistics and facilitating working with statistics for decision-making and teaching the methodology needed to produce appropriate statistics economically and to analyse and use them meaningfully.

Under pressure from the employers of statistical graduates such as government and non-governmental agencies, the universities in this region are attempting to reorient their courses towards application of statistics with some emphasis on practical aspects. This effort, though commendable, has generally been with uneven benefits mainly because of lack of teachers with background and experience in practical aspects of applied statistics. Further, skill in statistical operations such as evolving definitions, classification, designing of questionnaires, data collection, editing, coding, etc. can only be acquired through on-the-job training which is not possible in the universities due to lack of funds and facilities. Thus, if practical orientation of the university curriculum is to succeed, the facilities for exposure of the students to actual work should be provided through collaboration with statistical agencies.

Notwithstanding what has been said above, it must be noted, as mentioned by Rao (1979), that while the discussion of practical problems in teaching statistics was essential, it must not be done at the expense of the theoretical and mathematical contents of the courses. Though it is important that a statistician should be equipped with the practical skills to enable him to choose the appropriate methodology for analysis of given data, he should have at the same time the theoretical knowledge to formulate a new statistical problem for study if the existing methodology is inadequate in any given practical situation. Thus, what is needed is that we produce statisticians with the proper knowledge of theory and a flair for applications and give them opportunities to improve their skills during the course of their work.

Another aspect of statistical education in universities is the teaching of statistics to students of various disciplines, as statistics has an application in many fields. Here, the promotion of statistics can be ensured only if the subject is taught in such a way as to bring out forcefully the contributions of statistical knowledge to the development of the main subject of interest to the student which can be economics, demography, sociology, etc. Thus, practically oriented courses in statistics with a particular subject bias are needed in this situation instead of abstract theoretical courses.

The advent of computers and their accessibility to the university statistics departments are having a profound effect on the teaching of statistics in the universities. It is important that the universities should take account of the development in computer technology and capabilities and build this knowledge into the university curriculum. Many universities have already done this in this region and there is a keen desire to do so among the others also subject only to the necessary resources becoming available. Computer science is acting as a very good catalytic agent in galvanizing the development of statistical education in universities and this situation should be understood and utilised well in the interest of further development of statistical methodology and its applications.

REGIONAL TRAINING CENTRES

From what has been said in the previous section on university education in statistics, it is clear that the statistics graduates have generally adequate theoretical background in the subject, but the gap between statistical education and work requirements exists in spite of the effort to reorient the university curriculum towards applied aspects of statistics. This necessitated the creation and strengthening of facilities to provide the skills essential in statistical operations before the statistical graduates could be fruitfully employed in statistical work involving data production, analysis and interpretation, (ISI, 1951). According to the SIAP survey, though 10 out of 26 responding countries reported availability of national training facilities, 20 mentioned that the existing training facilities in the region were not adequate to remedy the situation, notwithstanding the fact that two international statistical training centres have been set up in this region to meet the training needs of the developing countries. They are the International Statistical Education Centre (ISEC) at Calcutta and the Statistical Institute for Asia and the Pacific (SIAP) at Tokyo.

The activities of ISEC are briefly discussed in ISEC (1977, 1982). The ISEC was set up in Calcutta in October 1950 for the training of statistical officers from the countries of the Middle, South and Far East. The Centre was sponsored by the International Statistical Institute and the Indian Statistical Institute under the auspices of and with financial support from the UNESCO. This arrangement was discontinued in 1952 but the work of the Centre was continued with the help of grants from the Government of India to the Indian Statistical Institute and also to the International Statistical Institute which made available teachers from abroad for the ISEC to supplement the faculty provided by the Indian Statistical Institute. The main purpose of the Centre is to provide courses in theoretical and applied statistics at various levels to selected participants.

The teaching at the Centre is mainly undertaken by the staff of the Indian Statistical Institute at Calcutta and this faculty strength is supplemented by visiting teachers provided by the International Statistical Institute, the United Nations and other international organizations. The Government of India collaborates with the Indian Statistical Institute in conducting the courses of ISEC by providing facilities for training in official statistics at the Central Statistical Organization and other ministries of the government.

Most of the trainees at the Centre are financially supported by the Government of India training grants under the technical co-operation scheme of the Colombo Plan, the Special Commonwealth African Assistance Plan (SCAAP) or other schemes. Some participants are under the fellowships provided by the United Nations and other international organizations. A number of trainees are also supported by their respective national governments or employer organizations. Originally the ISEC course was a six-month regular course, but some students were given the choice of continuing in the next term on some specialized topics. From 1954-55, the duration of the regular course was made nine months, but the choice of continuing in a subsequent term was no longer given to the participants. From 1966-67, the regular course has been made to be one of ten months' duration, thus the course length has been extended from the initial six months through the intermediate nine months to the present ten-month term. Also the Centre organizes special courses in collaboration with the CSO at places other than Calcutta but within India on specific topics, and these courses may be of varying duration.

The course design of the ten-month regular course consists of four phases of varying duration. During the first four months (June - September) at Calcutta, preparatory courses covering mathematics, practical work in probability and statistical theory and statistical projects such as sample surveys and data processing are conducted. This is accompanied by courses on economic statistics and national income statistics. During the second phase of six weeks (October - mid-November), the participants undergo training in official statistics organized by the Central Statistical Organization of the Government of India at New Delhi. The third phase covering mid-November to January and the fourth phase covering February - May are conducted at Calcutta. During the third phase of training, statistical methods, design of experiments, economics and economic statistics, sample surveys and demography are covered. During the fourth phase, each participant specializes in some field of study such as sample surveys, mechanical data processing, demography and vital statistics, statistical quality control and operations research, (ISEC, 1982).

The participants are evaluated on the basis of written tests and computations involved for analyzing data needed for the solution of practical problems. For the specialized training towards the end of the course, students are sometimes evaluated on short dissertations or reports on project work or through work assignments such as writing a computer programme. Candidates passing the final examination are awarded certificates of merit, while those who satisfactorily attend and complete the course but do not pass the final examination are awarded certificates of attendance.

For a detailed account of activities of SIAP, reference may be made to SIAP (1980) and Nair and Murthy (1971, 1973). The SIAP (formerly Asian Statistical Institute) was established in 1970 by the Governments of the countries of the Asian and Pacific region with the support of the United Nations Development Programme (UNDP) and the Government of Japan as an autonomous organization operating under the aegis of the Economic and Social Commission for Asia and the Pacific (SIAP). The Institute has completed two five-year phases and is presently entering its third five-year phase. The purpose of the Institute is to meet the training needs of the developing countries of this region in the field of statistics. At present the Institute conducts annually the following courses:

(i) General Course for six months for junior statisticians;

(ii) A three-month course on Automatic Data Processing for statisticians;

(iii) Two Advanced Courses or Seminars for four to six weeks on specialized topics for senior statisticians; and

(iv) Country or Sub-regional Courses on specific topics.

The fellowships for the General and ADP Courses are provided by the Government of Japan and those for the Advanced Courses and Seminars by the Government of The Netherlands. The faculty of the Institute is international in its character and is provided by the United Nations and the faculty strength is supplemented from time to time by experts made available to the Institute on non-reimbursable basis by host and other donor governments and international organizations. In the case of Country Courses the cost is generally shared by the Institute and the concerned countries.

The course design of the six-month General Course consists of a mandatory part and an elective part. The mandatory part generally covers sampling methods and elements of sampling, food and agricultural statistics, economic statistics, national accounting statistics, social and demographic statistics, statistical operations and automatic data processing. During the last three months of the course, in addition to some of the mandatory subjects, each participant specializes in one or two elective subjects. The subjects offered as electives vary from year to year, and in 1981-82, the two electives offered were sampling methods and demographic analysis. Greater choice of subjects is offered in some years depending on the availability of adequate faculty strength and interest among the participants. The course also includes field visits to observe various aspects of data production in Japan.

The emphasis in the General Course is on methodological, practical and applied aspects of statistics most relevant to the statistical work in government offices. Originally, it was a ten-month course during the first five-year phase and the duration of the course was reduced to six months since the beginning of the second five-year phase due to budgetary considerations, various demands on faculty time and other factors.

The fellows of the General Course, who generally have at least a basic degree, are first given diagnostic tests at the beginning of the course to evaluate their knowledge in basic mathematics and economic analysis and those who do not have adequate background in these two subjects are given supplementary courses on these two topics. In the case of the Advanced Course or Seminar, generally the fellows are senior statisticians with good academic background and considerable practical experience and hence the group turns out to be more homogeneous in their background and interest. Similarly, in the case of the Country Course on a specific topic, it is generally found that the background and experience of the fellows are basically of the same order. It may be mentioned that in an international course there is bound to be some amount of heterogeneity among the participants in respect of their educational background and work experience, and this should be kept in view in determining the course design and contents.

It is general practice in the Institute to provide the participants with a lecture outline or lecture notes and other relevant training documents well in advance of each lecture. Thus, over the last few years, the faculty members have prepared lectures notes in their areas, and these have been put in bound volumes for use in different courses. The notes generally include a brief description of the concepts, definitions, methods, operational procedures, illustrative examples, references and workshop assignments. These notes are kept in constant review in the case of the General Course and periodically updated using new data that become available. The Institute has produced lecture notes in the following subject areas: Statistical Methods, Sample Surveys, Agricultural Statistics, Demographic Statistics, Social Statistics, Economic Statistics, Economic Accounting Statistics, Statistical Operations, and Automatic Data Processing. Further, the documentation of the Advanced Courses are compiled together and kept in the library for the use of the faculty and the fellows. These lecture notes are available to the government statistical offices for their use in their training courses.

SIAP has recently initiated a UNDP-supported programme for the training of trainers of field enumerators and office processing staff. With financial support of the UNDP, SIAP collaborates with participating countries in producing relevant training manuals and training local trainers in the use of the training manuals and in training methods and techniques. The programme is underway in Indonesia and the Republic of Korea, and will eventually be expanded to include other interested countries. Its rationale is the belief that the quality of statistics produced is highly dependent on the motivation, knowledge and skill of field and office operatives.

In the case of the General Course, the participants' work is evaluated on the basis of a continuing system of written tests and assessment of completed workshop assignments. At the end of the course, the results of the evaluations are summarized in the form of consolidated scores on the basis of which certificates are issued to the participants. Those who satisfy the course requirements are given a certificate of successful completion and those who do not are given a certificate of attendance. Also, the fellows are given an opportunity to make their own assessment of the course through the use of a structured questionnaire. On the basis of the results of this enquiry, the course design and contents of the General Course are reviewed and wherever considered necessary changes, if any, are made for the next general course. This practice of evaluating the participants by the faculty and of the course by the participants is carried out in the case of the Country Courses also.

NATIONAL TRAINING CENTRES

As mentioned earlier, training centres have been necessitated by the gap that persists between statistical education and work requirements. National governments are trying to reduce this gap by creating and strengthening training facilities, which is generally a time-consuming and costly process. From the SIAP's survey, it is clear that only 3 out of 26 reporting countries felt that their training facilities were adequate for their purpose, while others expressed that their training facilities at national level can be realized when the rapid expansion of statistical services over the years is kept in view. Further, the two international statistical training centres in the region can at best meet only partially the training needs of the developing countries of the region.

The need for training at primary level is keenly felt whereas the international training centres generally aim at training statistical personnel at intermediate and professional levels. However, the training of trainers programme of SIAP tries to help this situation by training teachers to conduct training courses at primary level. Though many countries mentioned that their training facilities were inadequate, it is interesting to note that 10 out of the 26 countries in the SIAP survey reported creation of training facilities to train their staff. What is needed in these countries is to strengthen these facilities through proper training of teachers, supply of lecture notes, other training documents and teaching aids such as overhead projectors, slides, movies, etc. and rationalizing course design and content. In the case of countries where no facilities exist, the possibility of having basic courses to train primary and middle level personnel should be explored. The international centres can help the countries in establishing and strengthening their training facilities to serve their requirements at primary and intermediate levels. The problems of recruiting, training and retaining of statisticians in government service in the region were discussed at the Seminar on Statistical Organization held at Wellington, New Zealand, (Miller, 1979).

In the case of very small countries the possibility of developing a sub-regional training centre may also be explored for the purpose of conducting basic courses for sub-professional staff, and the requirements at professional level can be met through training in international regional centres.

TEACHING OF STATISTICS IN SCHOOLS

Of the 25 countries responding to the SIAP survey, 14 countries are already teaching statistics in schools and the other 11 countries are not doing so. In countries where statistics is taught in schools, it is generally introduced at the secondary school level, though a few countries have reported introducing graphical presentation at primary level. Further, the coverage is usually confined to descriptive statistics consisting of charts and graphs, frequency distributions and measures of central tendency and dispersion. However, a few countries are taking up some statistical analysis at the higher secondary school level covering correlation and regression analysis.

Though teaching of statistics in schools is of fairly recent origin, this has been strongly recommended much earlier by the Committee on Teaching of Statistics of the Institute of Mathematical Statistics (1948), Council of Royal Statistical Society (1947) and Mahalanobis (1951). ISI had discussed this topic at length at the Fourth Round Table Conference on Teaching of Statistics (Breny, 1976).

It may be stressed that if teaching of statistics in schools is to be taken seriously, standard textbooks would have to be developed through the organisation of task forces and these would have to be tried out in some selected schools with the help of trained teachers. Needless to say that the textbooks should be simple, illustrated with pictures and numerical examples and should bring out clearly the role of statistics in data production and inference. A concerted effort should be made to train the teachers through modification of the syllabus of the degree courses in education in the colleges and the universities. These points have also been brought out in the responses of some of the countries to the SIAP survey.

REPORTS FROM COUNTRIES

In the past, as part of its activities, the Statistical Institute for Asia and the Pacific (SIAP) has carried out surveys of training needs of the developing countries of the ESCAP region and of the extent of utilization by the countries of the training provided by the Institute through its various courses. These surveys had concentrated on the training aspects involved in producing an effective statistician as the Institute is basically a training institute, striving to provide the skills necessary for statistical work in the developing countries. Also, the Institute periodically tries to find out the specific requirements of each country in the region by correspondence or through the visits to selected countries by the faculty members as part of the activities of the Institute.

Recognizing that statistical education and statistical training are equally important to produce efficient statisticians, the Institute has been interested in both these fields to get a proper perspective of the statistical education and training facilities available in the developing countries of the region. For this purpose and to support the First International Conference on Teaching Statistics to be held at Sheffield, United Kingdom, during 8-13 August 1982, the Institute carried out a survey among the developing countries of this region. There is considerable interest in the outcome of this effort, as it would help in orienting its future programmes and plans to better meet the requirements of the countries of the region. Further, it is hoped that the result of the survey will provide a solid basis for further discussions of the problems of statistical education and training at national and international level in the region. For the purpose of the survey the following four forms were used:

(a) Statistical education at college level;
(b) Training in statistics;
(c) Statistical education at school level; and
(d) Overall comments and suggestions.

In the survey, statistical education at college and school levels and statistical training were separated to facilitate supply of required information, as the sources of information on them are usually different in the countries of the region. Before actually conducting the survey, the forms were first sent to India, Indonesia and the Phillipines for pretesting, which proved to be very useful. On the basis of this experience, the forms were finalised and mailed to all the developing countries of the region. Twenty-six countries have responded to the survey and only a few are yet to respond.

An overview of the results of the SIAP's survey is given in this section and detailed summaries of the reports from the countries so far received are given in Annex 2. In the table on statistical education and training facilities of Asia and the Pacific given in Annex 1, summary of the responses from the countries on the following are given: university facilities, computer facilities, demand for statisticians over supply, gap between education and work requirements, training facilities and teaching of statistics in schools.

If there is any basic degree course in statistics offered by any university in the country, the country is said to have a university facility. However, if the available facilities are considered insufficient to meet the requirements in the sense of lack of qualified teachers, number of students educated being less than the demand for statisticians, etc., the facilities are considered inadequate. If the educatioin facilities meet fully the requirements of the country in quality and quantity, the country is said to have adequate educational facilities. Similarly, in the case of computer and training facilities also the countries are classified as those where they are adequate and those where they are not adequate. In the case of a gap between education and work requirements, the countries are classified as those which are stating that the gap exists and those which do not. And for countries where no educational or training facilities exist this information is reported as not applicable.

From the SIAP survey, it emerged that 16 out of 26 responding countries had university facilities for teaching statistics and of these only 7 mentioned that these facilities were adequate. Of the 16 countries reporting university facilities, 14 mentioned that there was a gap between statistical education and work requirements necessitating further training of university graduates before their assignment to specific statistical work. Further, 21 out of the 26 countriers reported that the demand for statisticians was more than the supply with the available educational system. Regarding computer facilities, only 4 countries in the region reported that their facilities were adequate, whereas the others had mentioned that they either had no such facilities or that their facilities were inadequate and that they had to be strengthened.

Regarding training facilities, 10 of the 26 reporting countries mentioned that they had national training facilities in the field of statistics and the remaining 16 reported that they did not have formal training facilities. However, 23 of the reporting countries including 7 of the countries having training facilities reported that such facilities were inadequate. In fact, only 3 countries in the region, namely India, Hong Kong and Singapore, reported that they had adequate training facilities. In the case of teaching of statistics in schools, the 26 reporting countries were divided 15-11 between those teaching statistics in schools and those not teaching statistics at school level.

The general impression one gets from the SIAP survey is that statistical education and training facilities are gradually emerging in the countries of the region taking into account also the necessary computer facilities. In spite of the problem of not having trained and experienced faculty, lack of adequate teaching aids, books, etc., the countries have progressed in introducing statistics in graduate and post-graduate courses in the universities. The importance of statistical computing is well realized and appreciated and there is keen interest in expanding computer facilities in the universities in the context of statistics courses. Further, there is an increasing attempt to make the university courses in statistics oriented towards applied aspects of the subject. However this attempt has not fully succeeded due to various reasons including lack of teachers trained in practical aspects of statistical work. Thus, many countries reported that the gap between statistical education and work requirements exists and hence the countries have gradually tried to remedy this situation by creating and strengthening training facilities in their countries as well as in the region. Though statistics is taught in schools in many countries of the region, a number of problems regarding training of teachers, textbooks, etc. are yet to be discussed and resolved.

FUTURE DIRECTIONS

The complexities of modern economic and social life create a demand for information; and the advancement in computer technology and the availability of inexpensive, powerful computers make the creating of an adequate information base technically possible. Future needs therefore relate to our role as initiators, creators and users of the data base; that is, to:

(a) statistical awareness and motivation at policy levels in government and universities leading to the allocation of high priority and resources for the production and wide utilisation of an adequate statistical data base;

(b) development of the quantity and quality of skilled manpower needed for the production and utilisation of the data base; and

(c) education of users at all levels, particularly in government and business but also of the common people who need to understand and use quantified information for managing the affairs of ordinary life and even for intelligent understanding of news and events.

While (b) and (c) clearly give rise to needs for education and training, they are, at least in the short term, dependent on (a). If governments or universities do not recognize the central importance of quantitative thinking and statistics, they are unlikely to give adequate priority to statistical education in schools and universities. However, even a modest increase in the priority accorded to statistical education and training, particularly in schools, will eventually produce a more articulate and sophisticated demand for statistics and consequently an increase in priority and resources. Future efforts might therefore be concentrated in three main directions:

(i) increasing the statistical awareness and priority at policy levels in government;

(ii) widening of statistical curricula in universities to include practical aspects of applied statistics and promotion of the study of statistics; and

(iii) general statistical education in high schools.

More specifically, the following suggestions are made in respect of future needs in the field of statistical education and training in developing countries of Asia and the Pacific:

(i) initiative by national government statisticians to demonstrate the importance of statistics by making full use of existing statistics, particularly their analysis and presentation in forms which meet or partly meet the needs of official and other users. This inevitably increases awareness of the importance of statistics, sharpens the users' perceptions of their statistical requirements and leads to a higher priority being accorded to statistics;

(ii) Conduct of regular seminars by government statisticians and universities to introduce government and non-government users to the availability and uses of statistics and statistical methods;

(iii) Active promotion of statistics as a course of study and a career interest by official and other statisticians through campus interviews and promotional talks and the distribution of publicity brochures in universities and high schools;

(iv) Review of university curricula to reflect the practical aspects of applied statistics without diluting the conceptual and theoretical foundations necessary at the university stage;

(v) Revising the university curricula to include statistical computing as an essential part of the courses;

(vi) Training of teachers in universities and making computer specialists and practitioners available to universities to teach effectively statistical computing and practical aspects of applied statistics;

(vii) Achieving a balance in research work by encouraging research scholars to work on practical as well as theoretical problems, possibly through exchange of personnel between academic institutions and statistical offices of government, semi-government and other agencies;

(viii) Developing of national capability to train at least the personnel at primary and middle levels in statistical offices of government and other agencies;

(ix) Regional institutions such as SIAP helping the national efforts mentioned in (viii) through provision of training materials and aids in the form of lecture notes and structured training manuals, and with consultancy services;

(x) Encourage statisticians and statistical offices to work on methodological problems having relevance to present or future work of such offices and to publish their results in established journals and to attend seminars, professional meetings, etc.;

(xi) Collaboration between the United Nations Statistical Office, the United Nations Specialized Agencies, the International Statistical Institute and other relevant bodies in bringing the theoreticians and practitioners together for the development of statistics, and a much greater emphasis by all these agencies on the needs of the developing countries;

(xii) Professional journals reviewing and modifying their editorial policy to encourage publication of practical results of methodological interest and formulation of practical problems for possible solution by theoreticians, if necessary, through enlarging their editorial boards and referee panels to include practitioners and subject-matter specialists to ensure publication of papers with a practical orientation;

(xiii) Strengthening of computer capabilities of the universities and statistical offices with a view to utilising such facilities in teaching, training and practical work in an effective manner;

(xiv) Promotion of teaching of statistics in schools as a mandatory subject, at least, from the secondary school level;

(xv) Review and standardisation of textbooks and teaching methods for teaching statistics at school level;

(xvi) Organization of national and international panel discussions, seminars and conferences on various aspects of statistical education and training; and

(xvii) Special attention to be given to the countries in the Pacific region by strengthening the existing University of the Pacific at Suva, the Papua New Guinea University at Port Moresby, the University of Samoa at Apia and by creating sub-regional facilities for training of personnel at sub-professional level.

CONCLUDING REMARKS

Though it is made out in the paper based on the SIAP survey that there has been a gradual progress in the development of statistical education and training facilities in developing countries of Asia and the Pacific, it must be stressed that there is no room for any complacency, as the facilities created are far from adequate in many countries and such facilities are non-existent in some countries. It is expected that the national and international agencies concerned with statistical development in the region would be interested in the results of the SIAP survey and the views expressed in this paper. In the interest of statistical development, it is urged that these agencies should carefully consider the suggestions given in the section 'Future Directions' in this field with a view to taking up necessary steps to implement these suggestions as early as possible.

References

[1] American Statistical Association (1978) Conference on Transfer of Methodology between Academic and Government Statisticians; Washington, D.C.

[2] Bjerve, P.F. (1976) Presidential address presented to the fortieth session of the International Statistical Institute; Bulletin of the ISI, XLVI, (1), 41-48.

[3] Breny, H. (ed) (1976) The teaching of Statistics in Schools; Proceedings of the Fourth ISI Round Table Conference on Teaching of Statistics, The Hague.

[4] Central Statistical Organisation (1970) Statistical System in India; Department of Statistics, Government of India, New Delhi.

[5] Council of the Royal Statistical Society (1947) Report on the Teaching of statistics in universities and university colleges; Journal of the Royal Statistical Society, 110, 51-57.

[6] Daniels, H.E. 1975) Statistics in universities - a personal view: The Address of the President (with proceedings); Journal of Royal Statistical Society, Series A, 138, 1-17.

[7] Duncan, J.W. and Durbin, J. (1980) Report of the International Statistical Institute Committee on the Integration of Statistics; International Statistical Review, 48, 139-168.

[8] Fellegi, I.P. (1978) Some thoughts on training needed for government statistical work; Conference on Transfer of Methodology, 43-52, American Statistical Association, Washington, D.C.

[9] Finker, A.L. and Washabaugh, B. (1977) Education and training of domestic and foreign government statisticians in the United States; Bulletin of the International Statistical Institute, XLVII, (1), 447-469.

[10] Institute of Mathematical Statistics (1948) The teaching of statistics - Report of the Teaching of Statistics; Annals of Mathematical Statistics, 19, 95-115.

[11] International Statistical Education Centre (1977) Report on the Activities of the ISEC; Meeting of Directors of International Statistical Training Centres, Calcutta.

[12] International Statistical Education Centre (1982) Prospectus for the year 1982-83; Indian Statistical Institute, Calcutta.

[13] International Statistical Institute (1951) Statistical education programme; Report of the Statistical Education Committee; Bulletin of International Statistical Institute, XXXIII, (1), 116-137.

[14] International Statistical Institute (1969) Round Table Discussion on the University Teaching of Statistics in Developing countries; Review of the ISI, 37, (1), 86-98.

[15] Madhava, K.B. (1951) Problems of teaching statistics in South-East Asia; Bulletin of ISI, XXXIII, (V), 313-322.

[16] Mahalanobis, P.C. (1951) The role of mathematical statistics in secondary education; Bulletin of ISI, XXXIII, (V), 323-334.

[17] Malinvaud, E. (1978) University teaching for future official statisticians; Conference on Transfer of Methodology, 53-67, American Statistical Association, Washington, D.C.

[18] Miller, J.G. (1979) Status, recruitment, training and retention of statistical personnel in government office; mimeographed, Seminar on Statistical Organisation, ESCAP, STAT/SSO/7.

[19] Montenegro, T.H. (1973) Recruitment, preparation and status of statistical personnel; Inter-regional Seminar on Statistical Organisation, Ottawa, Canada.

[20] Murthy, M.N. and Khan, T.M. (eds) (1977) Proceedings of the Panel Discussion on Communication and Cooperation between Theoreticians and Practitioners; Occasional Papers No.6, Asian Statistical Institute, Tokyo.

[21] Murthy, M.N. (1978) Statistical education, training and research and their utilisation; Conference on Transfer of Methodology, 68-93, American Statistical Association, Washington, D.C.

[22] Nair, K.R. and Murthy, M.N. (1971) Asian Statistical Institute; Journal of the Society of Agricultural Statistics, XXIII, (2), 152-162.

[23] Nair, K.R. and Murthy, M.N. (1973) Evaluation of the courses and fellows' performance; Proceedings of the Meeting of the Directors of International Statistical Training Centres held at Szentendre, Hungary, 215-226.

[24] Rao, C.R. (1979) Perspectives in statistics; Presidential address at the 42nd session of the International Statistical Institute, Manila.

[25] Statistical Institute for Asia and the Pacific (1980) 10th Anniversary 1970-80; SIAP, Tokyo.

[26] United Nations (1947-51) Reports of the Sub-Commission on Statistical Sampling to the Statistical Commission; E/CN/, 3/37, 52, 83, 114, 140, United Nations, New York.

[27] United Nations (1949) An international programme of education and training in statistics; United Nations, E/CN/, 3/56, New York.

[28] United Nations (1981) Handbook of Statistical Organisation; Volume I, Studies in Methods, F, 28, New York.

[29] UNESCO (1974) A survey of statistics teaching in developing countries; Paris.

ANNEX 1

STATISTICAL EDUCATION AND TRAINING FACILITIES IN DEVELOPING COUNTRIES OF ASIA AND THE PACIFIC

Country [1]	University facilities		Adequacy of computer facilities	Demand of statisticians more than supply	Gap between education and work requirement	Training facilities		Teaching of statistics in schools
	Availability	Adequacy				Availability	Adequacy	
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Afghanistan	Y	N	N	Y	Y	N	N	N
Bangladesh	Y	Y	N	N	Y	Y	N	N
Brunei	N	N	N	Y	na	N	N	Y
China	Y	N	N	Y	Y	Y	N	N
Cook Islands	N	N	N	Y	na	N	N	N
Hong Kong	Y	Y	Y	N	Y	Y	Y	Y
India	Y	Y	Y	N	Y	Y	Y	Y
Indonesia	Y	N	N	Y	Y	Y	N	Y
Kiribati	N	N	N	Y	na	N	N	N
Korea	Y	Y	Y	Y	N	N	N	Y
Malaysia	Y	N	N	Y	Y	N	N	Y
Maldives	N	N	N	Y	na	N	N	N
Nauru	N	N	N	Y	na	N	N	Y
Nepal	Y	N	N	Y	Y	N	N	Y
Niue	N	N	N	Y	na	N	N	N
Pakistan	Y	Y	N	N	Y	N	N	N
Philippines	Y	N	N	Y	Y	Y	N	Y
Singapore	Y	N	Y	Y	na	Y	Y	Y
Solomon Islands	N	N	N	Y	Y	N	N	Y
Sri Lanka	Y	N	N	Y	Y	Y	N	N
Thailand	Y	Y	N	Y	na	Y	N	Y
Tonga	N	N	N	Y	na	N	N	Y
Tuvalu	N	N	N	Y	na	N	N	N
Vanuatu	N	N	N	Y	Y	N	N	Y
Viet Nam	Y	Y	N	N	N	Y	N	N
Western Samoa	Y	N	N	Y	Y	N	N	Y
SUMMARY	Y-16 N-10	Y-7 N-19	Y-6 N-22	Y-21 N-5	Y-14 N-2	Y-10 N-16	Y-3 N-23	Y-15 N-11

[1] responding to the SIAP survey.

Abbreviations: Y - Yes; N - No; na - not applicable.

ANNEX 2

SUMMARY OF REPORTS FROM COUNTRIES RESPONDING TO THE SIAP SURVEY

1. Afghanistan

The University of Kabul provides a B.Stat. course for twenty candidates per year, emphasizing equally theoretical and practical aspects of statistics. Statistics is also taught in B.A.(Econ.) course. The university courses are supplemented by the training opportunities offered by the international agencies in and outside the country. Most of the graduates are absorbed in the government agencies. There is a shortage of professionally trained statisticians in the country, as there are no regular training facilities. The facilities currently available in the universities are also limited and minimal.

There is an informal on-the-job training carried out by the employees in the data processing in the Central Statistical office (CSO) and once a year instruction in data collection is offered to the provincial personnel. The CSO considers that there is a gap between statistical education and work requirements of a statistician and to reduce this gap the CSO is in consultation with governmental and academic organizations and would support them qualitatively and quantitatively in ensuring that practical procedures are introduced in the curriculum. There is a plan to establish an Administrative Training Centre to meet the training requirements of government personnel.

It may be mentioned that the SIAP has conducted two courses in Afghanistan, one is a workshop on basic aspects of general applied statistics in 1974, and a country course in demographic and social statistics with special reference to the population census in 1976. The country expects to have more contributions from SIAP in reducing the gap that exists between supply of and demand for statisticians.

2. Bangladesh

The University of Dacca offers graduate and post-graduate courses in statistics in the country. As part of the University, the Institute of Statistical Research and Training (ISRT) is participating in statistical education, training and research and it brings out the Journal of Statistical Research. There is the Bangladesh Statistical Association headquartered at the ISRT in Dacca. There is a training section in the Statistics Division in the Government of Bangladesh to co-ordinate the training activities in the country and many training courses have been held to impart specific skills to statistical personnel at different levels. SIAP has conducted training courses on sample surveys in 1977, management of statistical operations in 1979 and national accounting statistics in 1980. Also, the country sends its statistical personnel abroad for training in statistics. Bangladesh has hosted two FAO regional workshops on agricultural census and agricultural statistics in 1978 and 1981.

3. Brunei

There are no statistical education and training facilities in the country. Students wishing to take up degree or diploma courses have now to go abroad and most of them go to the United Kingdom for statistical education. Presently the demand for statisticians exceeds the supply. Basic elements of statistics are introduced in schools right from the lower primary classes. In the secondary forms, many subjects such as science and economics include statistics as part of the curriculum.

4. China

The People's University of China and the Universities of Xiamen, Liaoning and Hebei, the finance and economic colleges and the finance and trade colleges have been offering many courses in statistics in the country. At the bachelor's level, the most commonly offered courses are principles of statistics, mathematical statistics, industrial statistics, agricultural statistics and commercial statistics.

The facilities for statistical education and training available in China are relatively less developed and in most cases the faculty strength and capabilities are limited. computer facilities are available only in a few universities and colleges and the concerned personnel lack the necessary experience in handling them. There is shortage of trained statistical personnel in the country.

There is considerable interest in learning statistics, the modern technology, by increasing education and training facilities, sending appropriate personnel abroad, inviting experts to give lectures and conduct seminars and courses, and procuring books. The present aim is to train the teaching and professional staff concerned with statistical work by stages and in groups.

There is a plan to set up a planning and statistical institute and extend statistical institutions, which already exist at present at some places, to more places in the country. It would be useful if it is possible to train in about five years' time the present statistical personnel who do not have proper background in statistics. At present statistics is not being taught at school level in China.

5. Cook Islands

The country offers correspondence course in applied statistics with the co-operation of the University of the South Pacific in its diploma programme. Apart from this, the country has no statistical education or training facilities at present and depends to a large extent on external assistance on developing its statistical personnel.

6. Hong Kong

The University of Hong Kong and the Chinese University of Hong Kong provide graduate and post-graduate courses in statistics and the Hong Kong Polytechnic offers a diploma course in statistics. The bachelor's degree course is for three years at the University of Hong Kong and for four years at the Chinese University of Hong Kong. The master's degree course is for two to four years. In some of the courses, computer science is offered as a minor. Computer facilities available are adequate for purposes of statistical education and training. The Hong Kong Statistical Society was founded in 1977 and has a present membership of about 200.

The gap that exists between statistical education and statistical work requirements is being reduced by organizing training for the new recruits at the Census and Statistics Department, applying job rotation system, sending appropriate personnel abroad for specialized courses and attachment of personnel to statistical offices of developed countries. Statistics is taught in the secondary schools covering graphical presentation and measures of central tendency in first to third year and covering measures of dispersion of standard distribution and linear regression in fourth to seventh year.

7. India

India has about 50 universities offering statistics at graduate and post-graduate levels with the bachelor's degree being for three to four years and the master's degree being for two years. The Indian Statistical Institute, established in 1931, and the Indian Agricultural Statistics Research Institute founded in 1947 also offer graduate and post-graduate courses in statistics and carry out research work in both mathematical statistics and applied statistics.

The country has more than adequate facilities for teaching statistics and guiding research in this field. Though the curriculum in most of the courses concentrates on mathematical statistics, attempts are being made to make the courses include some emphasis on statistical operations and applications. Mechanical and electronic desk calculating machines are usually available for use in classrooms and computer facilities are increasingly becoming available. Many foreign students study statistics in Indian universities and institutes. The local demand for statisticians is adequately being met by the present educational system. In fact, there is some problem of unemployment among the graduate statisticians in the country.

In spite of the attempts to make the graduate and post-graduate courses oriented towards practical aspects of statistics, there is still a gap between statistical education and professional work requirements in this field. Training facilities have been created in the two institutes mentioned above and in the Central Statistical Organization (CSO) to bridge this gap. The new recruits in the statistical service of the country are given intensive training for two years before they are assigned to regular jobs in the government. Appropriate personnel are sent abroad for further training in case of specialised fields. SIAP participated in the training seminar on use of sampling in census held at New Delhi in 1980.

Recognizing the need for practical orientation in teaching statistics, some universities send their students to the CSO to study official statistics as part of their curriculum. The International Statistical Education Centre, which runs a long-term statistical training course for participants from various countries, is being supported by the Indian Statistical Institute, the Government of India and the International Statistical Institute. The fellowships for this course are open to candidates from any country in the world. The activities of this Centre are covered in greater detail in the section 'Regional Training Centres' in this paper.

Statistics is being taught in higher secondary and high schools in India. Basically, graphic representation is covered in the first five years of the school's curriculum and in the remaining years collection and compilation of data and descriptive statistics are covered.

The Indian Statistical Institute Association established in 1931, the Indian Society of Agricultural Statistics established in 1947, and the Calcutta Statistical Association founded in 1945 are well known statistical professional societies in India and they bring out Sankhyā, the Indian Journal of Statistics, Journal of the Indian Society of Agricultural Statistics and the Bulletin of the Calcutta Statistical Association, as part of their professional activities in statistics.

8. Indonesia

Statistics courses are offered by the Academy of Statistics (Jakarta), Institute of Technology (Bandung) and some universities. There is a continuing attempt to improve the facilities by having new classrooms, more teaching staff at the M.A. and Ph.D. levels, library expansion and acquiring of more computer capability. Further, the existing staff capability is also proposed to be upgraded through further education abroad.

The gap that exists between statistical education and work requirements in statistics is being bridged to some extent by creating training facilities at the Statistical Education Training Centre which conducts many training courses including some correspondence courses. There is a plan to upgrade and expand the training facilities by establishing training centres in some provinces. The country is participating in the SIAP's training of trainers programme. SIAP has conducted training courses on national accounts in 1975 and on distributive and service trade statistics in 1978 and has participated in the workshop on sampling in 1980. Statistics is taught at high school level in the country.

The Indonesian Statistical Society is headquartered at the Central Bureau of Statistics in Jakarta.

9. Kiribati

The country utilizes the University of the Pacific branch in Tarawa, which conducts a course on elementary statistics for six months. The country has also participated in the course on statistical operations and procedures of the south Pacific Commission conducted for six weeks in 1979. It is felt that one graduate statistician is badly needed to organize and carry out the statistical functions in the country. Statistics is being taught at the secondary school level in Kiribati.

10. Korea

Graduate and post-graduate courses in statistics are offered by universities in Korea. The bachelor's degree courses are usually for three to four years whereas the M.A. degree courses are for two years. The universities generally have adequate computer facilities for use in their courses. The demand for statisticians in the country is about 20% more than the supply available from the present educational system. Most of the courses in the universities have been oriented to applications of statistics to a large extent and this has helped in reducing the gap that generally exists between statistical education and work requirements in this field. However, even with this reorientation of the courses to practical applications, there is some gap that is felt and to reduce this gap it is planned to set up a centre for statistical survey and its analysis at Donaook University. Further, there is also a plan to have larger computers in the universities and also to expand the available educational facilities by about 20%.

SIAP has conducted short-term training courses at Seoul on National Accounts in 1972 and on Statistical Methods with emphasis on practical aspects in 1979. The country is participating in the SIAP's programme of training of trainers.

Statistics is taught at school level in Korea, the coverage being statistical tables, histograms, pictographs, arithmetic mean and frequency distribution at the primary level; relative frequency, cumulative frequency, meaning of sample and population at the junior high school level; and mean, standard deviation, frequency distributions and normal distribution at the high school level.

The country has a Statistical Association of Korea headquartered at the Bureau of Statistics at Seoul.

11. Malaysia

The University of Malaya, the National University and the Agricultural University offer courses in statistics at graduate level. There is a great need for statisticians with an economics orientation and attempts are made to make the statistical courses application-oriented in the universities. The computer facilities available at present are not adequate. The demand for statisticians exceeds supply in the case of applied statisticians and the situation is the reverse in the case of theoretical statisticians. The country has no training facilities as such at present. It is felt that there is considerable need to upgrade the skill of teaching staff through exposure to applied aspects of statistics at SIAP and other institutions abroad.

Statistics is taught at the GCE O-level in schools. However, the present syllabus in high school certificate course in the field of statistics is too advanced and proper training for teachers is necessary if these courses are to succeed.

12. Maldives

The country has no university or college. The statistical education requirements of the country are presently being satisfied by sending appropriate personnel abroad for this purpose. The shortage of statisticians is being keenly felt in the context of the development of the country's economy and social life. The country has no regular training facilities and is sending suitable candidates for training at international institutes such as SIAP at Tokyo, ISEC at Calcutta and ISPC (International Statistical Programme Centre) in Washington D.C. Recently, the UNDP expert on statistical development conducted an elementary basic statistical course on collection of data for about three months at Malé. There is a need to meet the requirements for statistical training personnel through training abroad at the present stage.

13. Nauru

At present the country has no statistical education and training facilities. As the population of the country is rather small, the scope for statistics is also limited. However, rudiments of statistics are taught in secondary schools as part of mathematics curriculum. Nauru is also a member of the University of the South Pacific, Suva (Fiji), where short-term and degree courses in statistics are offered to the nationals of member countries. But so far Nauru could not avail of this facility.

14. Nepal

The Tribhuvan University offers bachelor's and master's degree courses in statistics. Except for calculating machines, there are no computer facilities readily available for use in statistical education at university level. Since the

60

university courses are theoretical in nature, and since applied courses in statistics are not available, there is a gap between statistical education and work requirements. The demand for statistics still exceeds the supply, though the position at present is better than before.

There are no training facilities in the country in the field of statistics. A few agencies in the government such as the Central Bureau of Statistics have benefited from getting some of their staff trained at institutions such as ISEC and SIAP. However, such opportunities are not easily available to the educational institutions to upgrade the quality of their staff. It is to be noted that the success of developmental planning in future will depend on the upgrading of the skills of the producers and the users of statistics in the country.

Elementary ideas of statistics are taught in schools from the eighth grade onwards.

16. Pakistan

The universities in Pakistan offer graduate and post-graduate courses in statistics and the facilities for teaching statistics are quite adequate and there is not much imbalance between supply and demand of statistics in the country. Pakistan Statistical Association is headquartered at the Institute of Statistics in the University of Punjab at Lahore.

There is a gap between statistical education and work requirements and this gap is being bridged through orientation and introductory courses conducted by the Statistics Division for the statistical personnel. However, at present, there is no institutional set-up for training as such in the country. The Statistics Division has training section and it is expected that the role of this section would be expanded to facilitate training of the statistical personnel to enable them to perform their duties effectively. There will still be the need of foreign facilities to upgrade the capabilities of the statistical personnel in specialized fields. In future it is expected that the direct recruitment at middle level in the statistical cadre would be through competitive examinations covering both theory and practice of statistics. The country presently utilizes the available fellowships under various plans such as the Colombo Plan, UNDP programmes, SIAP, ESCAP and East-West Centre, but it is found that these facilities are still not adequate to meet the growing requirements within the country. SIAP has conducted short-term training courses at Karachi on Sample Surveys in 1979 and on National Accounts in 1981.

Statistics is not being taught in schools in the country.

17. Philippines

The University of the Philippines and its Los Banos Center, the Polytechnic University and the University of the East offer graduate and post-graduate courses in the field of statistics. The graduate courses have been oriented towards practical aspects to the extent feasible. The aims of these courses are to produce professional statisticians and teachers and research workers. In spite of these facilities, there is a serious shortage of qualified statisticians. Appropriate statistical personnel are sent abroad for education and training, but these facilities are limited and are not sufficient to meet the growing demand for technical expertise in the country. The University of the Philippines at Los Banos is planning to include use of computers in its courses. It also has programmes to develop a Ph.D. course and statistical computing facilities. Similarly, the University of the Philippines is planning to revise its curriculum to reflect the specialists' needs from its present generalist orientation.

There is a gap between statistical education and work requirements in statistics and to bridge this gap to some extent training facilities are being created at the University of the Philippines at Los Banos, Statistical Centre of the University of the Philippines National Census and Statistics Office and National Economic Development Authority (NEDA). The NCSO provides on-the-job training for non-statisticians and also conducts a training course on EDP. Also, NEDA conducts non-degree, work-orientation courses for its staff.

A national convention of statisticians is held biennially to exchange views and discuss problems among professional statisticians in the country. There is an imbalance between supply and demand of statisticians in the country as about 48% of the staff have had no formal training in statistics. The NCSO is undertaking a continuing statistical training programme to develop various statistical training models for different levels of statisticians to assist other government agencies in the development of statistical methodology, to set up an effective method of administration of statistical training and staff development by providing and expanding scope of statistical training to its personnel in co-operation with the universities and other agencies in the country and abroad. It is also proposed to develop and maintain a system of evaluating the results of such efforts. The Statistical Centre of the University of the Philippines is proposing to have a department of training and administrative services to provide continuing education for statistical centre alumni, training for statistics teachers and consultancy services for government agencies.

The NEDA is considering the possibility of setting up a statistical institute for the Philippines which would conduct professional degree and non-degree courses in applied statistics and also organize seminars and workshops in the specialized fields in statistics for improving statistical knowledge and upgrading skills of the appropriate personnel. The institute will also develop training materials, conduct studies on more effective methods for conventional and specialised training in statistical work and carry out research on recent developments in applied statistical methods, data processing and analytical techniques.

It is felt that SIAP should increase training programmes and fellowships for trainees on the advanced courses in statistics and in computer sciences. SIAP has conducted a short-term course on National Accounts in the country in 1977 and has participated in a data processing course of the National Census and Statistics Office in 1981. The United Nations and its agencies should help developing countries acquire hardware for computer usage.

Statistics is taught in schools. Elementary statistics is an elective subject in the fourth year of secondary school courses but only a few schools offer this and very few students take up this option.

The Philippines Statistical Association has a membership of about 350 and brings out the periodicals Philippines Statistical and PSA newsletter.

18. Singapore

The National University of Singapore offers bachelor's degree course emphasizing both theory and practice of statistics. The course is of four years' duration and trains sufficient number of candidates for supplying the local demand. The universities have adequate computer facilities but however in practice little training is possible on the computer itself. There is no difficulty in graduate employment in the country. It is expected that in the near future the statistics education facilities will be increased by about 20 per cent. Also statistics is taught as a subsidiary subject in many other courses with an orientation towards applications.

In spite of the practical orientation of the university courses, there is a gap between statistical education and work requirements and to reduce this gap short orientation courses are given to the new recruits and also appropriate personnel are sent abroad for further training or practical attachment with statistical offices of developed countries in specialized fields. With the latest emphasis on computerization in the civil service, statistical personnel will also be trained to handle computer equipment and operate micro- and mini-computers to upgrade the operations of calculation, processing and compilation of statistics. There is still a shortage of trained statisticians at graduate level as many statistical graduates find employment as economists or administrators.

Statistics is taught at the secondary school level and no difficulties are experienced in this effort.

19. Solomon Islands

The Solomon Islands Extension Centre of the University of the South Pacific offers a course in applied statistics as part of its diploma in accounting studies. The country has no other course in statistics and the present arrangement with the University of the South Pacific is considered adequate. The country has participated in the SPC's course on statistical operations and procedures (SOAP), which is a six-week course with practical orientation. There is a specific need in the country for sub-professional training and long-term overseas courses are not feasible as the number of personnel involved in the statistical office is very few. It is expected that some candidates will make use of the Institute of Statisticians' Examination offered by correspondence courses. It is desirable to supplement the SPC's SOAP course by correspondence or extension courses in applied statistics.

Statistics forms part of mathematics and social science syllabus in Solomon Islands' school certificate taken up to five years of secondary education and the emphasis is on basic numeracy and using statistics in real life situations.

20. Sri Lanka

The Universities of Colombo, Jaffna and Jayawardanapura offer graduate and post-graduate courses in statistics, emphasizing both theoretical aspects, the bachelor's degree course being for three years and the master's degree course being for one to one and a half years and the diploma post-graduate course being for one year. Statistics is also taught as a subsidiary subject in a number of other courses. The faculty generally faces shortage of teaching expertise and computer facilities in conducting the courses. The need to introduce automatic data processing as part of statistics curriculum is being increasingly felt. There does not seem to be any serious imbalance between supply and demand of statisticians in the country. There is also need for strengthening the statistics faculty in the universities and also to make available better computer facilities to carry out the statistics courses more effectively. The University of Colombo has link arrangement with the Department of Applied Statistics of the University of Reading and such arrangements are considered to be very beneficial for upgrading the teaching capability of the faculty in the universities.

The gap that exists between statistical education and work requirements is keenly felt and the Training Division in the Department of Statistics conducts occasional training courses at different levels to impart the necessary skills to its staff. The training facilities locally available are not adequate to meet the growing demands and foreign training facilities at SIAP, East-West Centre, ISEC and ISPC are being utilized to meet the demand for trained statisticians. SIAP has conducted short-term courses in Colombo on sample surveys and statistical indicators in 1975, agricultural statistics in 1977, demographic and social statistics with special reference to population census in 1978 and national accounts in 1979.

Statistics is not being taught in the schools in the country.

Sri Lanka Statistical Society has its headquarters at the Department of Census and Statistics in Colombo.

21. Thailand

The universities in Thailand offer graduate and post-graduate courses in statistics with a mix of theory and practice, the bachelor's degree course being for four years and master's degree course being for two years. Desk calculating machines and many computers are generally available for use in the courses. There is not much of an imbalance between supply and demand of statisticians in the country. It is expected to revise the course contents in statistics with a view to orient it towards applications of statistics. Also it is intended to enlarge the computer facilities.

In spite of the practical orientation of the university courses in statistics, there is a gap between statistical education and work requirements, as students in colleges have little working knowledge of statistical operations as carried out in real-life situations. The National Statistical Office organizes training programmes for four to six months to reduce the gap between statistical education and work requirements by imparting statistical skills to its staff. Thus, the balance between supply and demand of statisticians is only in quantity but not in quality as in practice the graduates do require further training to make them useful in statistical offices. The country also utilises training facilities available abroad. SIAP has conducted courses on evaluation analysis and use of statistical data in 1978 and on assessment and control of errors in 1981.

Statistics is taught at school level in the country covering interpretation of graphs and charts at the primary level, elementary basic statistics at the secondary levels and elements of statistical theory at the high school level.

Thailand's Statistical Association is headquartered at the National Statistical Office at Bangkok.

22. Tonga

The country has no statistical education and training facilities and relies mainly on assistance from abroad to meet its statistical personnel requirements. The country has had the SPC's SCAP course for six weeks and also specialised country course for middle-level personnel for four weeks. Calculating machines are generally available for use in courses and in work. The shortage of trained personnel is particularly felt at middle-level and there is need for organizing primary and middle-level training programmes to upgrade the skills of the statistical personnel.

Statistics is taught in schools from form four level and it is expected to introduce more advanced statistics at form seven level.

23. Tuvalu

The country has no statistical education and training facilities and there are no plans for including statistics in the school curriculum. The present requirements of statistical personnel are met through external assistance in this field.

24. Vanuatu

The country has no statistical education and training facilities. With a population of 113,000 the country has limited scope for statistics. The country has utilised the SPC's courses on external and internal migration and the statistical operations and procedures. Calculating machines are available for use in the courses and in work. The country has three experts in the field of statistics who are expatriate staff, and it is expected that this staff will conduct training for the local staff at sub-professional level. For the purpose of professional training, it may be necessary to use the facilities available abroad.

Statistics is taught in schools with simple charts and statistical measures of central tendency being covered in three years of post-primary education and statistics and probability being covered at the GCE O-level.

25. Vietnam

The Economics Planning Universities in Hanoi and Ho-Chi-Minh City conduct economic statistics courses for four years with emphasis on practical aspects. Statistics is also taught as part of the book-keeping courses of the central statistical cadre and secondary statistical school. The emphasis in these courses is on statistical survey methods and reporting of forms. There is no gap between supply and demand as supply is planned on the basis of the requirements of the government and other agencies. Specialized courses on statistics balance and on sample surveys have been conducted in the country. For higher education and research, appropriate personnel are sent abroad. The country expects to benefit from the course design and course contents used at SIAP.

Statistics is not taught in schools in the country.

26. Western Samoa

The Extension Center of the University of the South Pacific at Apia and the University of Samoa offer five-year bachelor's degree courses in statistics and mathematics covering only the basic aspects of the subjects. There is a serious imbalance between supply and demand of graduate statisticians in the country. The practising statisticians generally receive the necessary training in statistics overseas in institutions such as the Statistical Institute for Asia and the Pacific at Tokyo, the International Institute for Population Studies in Bombay and the East-West Center at Hawaii. The training facilities offered by ESCAP and SPC are also utilised.

The primary and secondary teachers colleges offer courses in mathematics covering elementary and basic aspects of statistics in their short-term courses, which are oriented towards primary and secondary school teachers. Statistics is taught in schools in the country covering basic elementary aspects of statistics in form 2 as part of mathematics syllabus and from form 3 statistics is taught as a part of applied mathematics.

It is suggested that the policy-makers should undergo an appreciation course in statistics, the teachers and lecturers should have intensive/university education and the practicing statisticians should have in-service training in some recognised statistical offices after completing statistical training abroad.

COMMENTS ON THE ABOVE PAPERS

K T de Graft Johnson, Economic Commission for Africa, Addis Ababa

The three papers give brief descriptions of the development of statistical education and training in Africa, Asia and the Pacific and Latin America and discuss problems created by the gap between theory and practice which in spite of attempts to bridge it still persists to this day. Before commenting on general issues, I will make some remarks on each paper separately.

Adichie and Afonja state that what is taught in statistical institutions is dictated by the priorities of the country and the university. I doubt whether that reflects accurately what is actually happening in most of the African universities and training centres. For example, energy is currently the most important element in the development of the countries of the region, whether these countries are oil-producing or not, but how many of these educational and training institutions include a course on energy statistics in their curriculum? The truth is that most of the courses are carbon copies of what is taught in British and American universities. It is only recently that conscious efforts are being made at least in the field of applications to shift the focus to be in line with national priorities.

The authors also refer to the admission of statistics as a discipline in its own right as a recent event. This is true of Africa but the underlying reasons are rather complex. While, for example, the University Colleges of Ibadan and the Gold Coast were first established with predominantly British academic staff, it was only natural that they should follow the Oxford and Cambridge rather than, say, the Iowa State tradition. The majority of students in the initial stages were general rather than special or honours degree potential. My own graduating class in 1953 had only two candidates for the special degree in mathematics; and of course there were only three questions out of ten in one of the eight papers required for the B.Sc. (Special in Mathematics) examination of the University of London which related to statistics. You could therefore take the whole examination without answering a single question in statistics. It may be recalled that during the period under reference, the general degree was of a two years' duration while the honours degree lasted three years, both after the intermediate examination. As seniority in the civil service was based on year of entry, honours or special degree holders had the disadvantage of working under their classmates who had opted for the general degree. My own recollection of the first person in the then Gold Coast (now Ghana) who could be described as a professional statistician having done a degree in mathematics and also passed the Institute of Statisticians' examinations with honours, was that he was refused entry into the statistical office because he was too specialised to be useful to the Central Bureau of Statistics. Thus job openings and the academic inheritance of the African universities were the main problems encountered in statistical training in the initial stages in English-speaking Africa. The position of the French-speaking countries was only slightly better, since they were trained at the CESD, but it should be noted that this was modelled on the lines of the French ENSAE.

Adichie and Afonja also list two major constraints in statistical education, namely, inadequacy of qualified teachers and non-availability of teaching material in applied areas. However, as implied in my preceding remarks the position is more complex.

I next turn to the paper by Cansado. He refers to the problem of finding suitable staff, namely persons with a good theoretical and practical background for the statistical training centres. This is a serious problem for all training centres in the developing world. I will revert to this issue in my concluding remarks.

Cansado also states "A characteristic of university teaching in this part of the world..... is that in the formal teaching a heavy weight is given to statistical theory and a very light one to applications and actual data analysis". The question is "need this be so?" It is my view, for example, that in teaching sampling or design of experiments there is great scope for illustrating the theory by reference to applications and actual data analysis. If this is done, most of the criticisms directed against the so-called academic statisticians by the government statisticians would no longer be valid.

Another valid point made by Cansado is the very little research done in Latin American Universities. He quotes the recommendations of a 1980 Working Group on this subject which I hope will be discussed extensively here. There is one question which should be addressed to all authors of papers for this session in connection with the recommendations. How much, and what type of research, is any developing area government willing to pay for? My experience in a number of developing countries has been that governments are willing to pay very little for pure or applied research. Most research projects have thus to be camouflaged by emphasising the immediate policy relevance of the results rather than the wider and long-term benefits.

I will now deal with only two issues in the paper by Murthy; namely future directions and his suggestions for action. He emphasises three main areas in which actions should be taken:

"(i) increasing the statistical awareness and priority at policy levels in government;

(ii) widening of statistical curricula in universities to include practical aspects of applied statistics and promotion of the study of statistics;

and

(iii) general statistical edcuation in high schools."

These recommendations are generally in line with those which have been made by various working groups and meetings of directors of African statistical training centres and approved by the African Conferences of Ministers of Planning. In fact the region has gone even further to recommend that national statistical development programmes should be included as part of national development plans. This view although accepted by the OAU meeting of heads of state has yet to be widely implemented. This is an example of a different type of gap - between resolutions and their implementation.

As stated earlier, I will now deal with general issues arising out of the three papers. These relate to the implied dichotomies between education and training, and academic vs government statisticians as well as the role of research in statistical development.

My dictionary defines training as "bringing a person to the desired state or standard of efficiency, etc., by instruction and practice", and education as "giving intellectual and moral training to a person". Thus as far as statistical development is concerned there should have been considerable overlap between education and training. Failure to expand this overlapping area is the cause of the present unsatisfactory situation. It is worth noting that in one African University, the Institute of Statistics and Applied Economics at Makerere University has attempted to expand this area of intersection between training and education by introducing a degree course B.STAT in addition to the traditional B.Sc.Stat. The former deals extensively with data handling techniques while the latter only briefly considers official statistics such as agricultural and industrial statistics. In line with my own views on the matter, I would have been much happier if the institute could have evolved only one degree course for statistical specialists which brought together the best elements in the B.STAT and B.Sc. courses.

Implicit in all three papers is the rather erroneous assumption that there are only two groups of statisticians - academic and government. The ISI Committee on Integration of Statistics identified 15 categories, namely:

Academic	(including university, college, school)
Business	(including actuarial, marketing, economic, econometrics)
Government	(including federal, state, local, municipal)
Industrial	(including operations researcher, quality controller, reliability worker, physical, engineering, transport)
Practitioner	(including applied statistics, data analyst)
Research	(including applied statistics, methodological, theoretical, mathematical, probabilist)
Student	
Biometric	(including ecological, agricultural, forestry)
Medical	(including biostatistician, health, epidemiologist, pharmaceutical)
Social	(including demographic, political, judicial, criminal, behavioral, psychological)
Administrator	(including manager, policy former, planner, forecaster, auditor)
Consulting	
Computational	(including simulator, computer programmer, package writer)
Survey	(including pollster)
Other	(including sports, numerologist)

In my view, it is more relevant and useful to deal with the content of statistical training in developing countries in the broader spectrum of 15 different categories than in the narrower context of a simple dichotomy of academic and government statisticians.

At this stage, it is worth discussing the main recommendations of a 1981 Working Group on Statistical Organization convened by the United Nations Economic Commission for Africa which considered inter alia, the integration of statistics and its implications for teaching. The meeting "considered that integration was desirable but that it could be achieved more effectively by improvements in the methods of training which produced a proper mix of theory, methods and applications. ..... Stress was laid on the period of apprenticeship either as part of a degree programme or outside it".

In conclusion, Mr. Chairman, I would like to refer to the need for research - both substantive and methodological - by all 15 categories of statisticians in developing countries. This emphasis is implied in Cansado's paper and I endorse it whole-heartedly. Research can also bridge the gap between the academic and government statisticians, if both engage in it. In addition those who teach official statistics and data handling techniques in universities should have had considerable practical experience in these fields. This regrettably is not the case in most developing countries.

REPORT ON SESSION ON STATISTICAL EDUCATION AND TRAINING IN DEVELOPING COUNTRIES

B P Adhikari, Indian Statistical Institute

The session consisted of a paper reading meeting and two workshops.

At the paper reading meeting 3 papers were presented:

1. Statistical Education and Training in Latin America - Enrique Cansado, CIENES, Santiago, Chile.
2. Statistical Education and Training in the Developing Countries of Africa - J.N. Adichie, Department of Statistics, University of Nigeria, Nsukka and Biyi Afonja, Department of Statistics, University of Ibadan, Ibadan.

and

3. Statistical Education and Training in Developing Countries of Asia and the Pacific - M.N. Murthy, Statistical Institute for Asia and the Pacific, Tokyo.

The first two papers were presented by the authors themselves. Since Dr. Murthy was not present, the salient points made in his paper were presented by the Chairman. Copies of all three papers had been sent in advance to Dr. K.T. de Graft-Johnson of the Statistics Division, Economic Commission for Africa, Addis Ababa and he presented his comments. There was an open discussion on the papers which, together, gave a broad picture over all the developing countries of the world. Since copies of the paper had not been made available in advance to participants and the abstracts too brief, it was difficult for them to focus on specific issues raised in the papers. Participants from several countries disagreed with some of the findings of the survey reported in Dr. Murthy's paper. Critical comments on his paper were made particularly by participants from India and the South Pacific region.

The two workshops partly dealt with the contents of the three papers and partly on other issues raised on the floor. Some of the points made were:

1. The level and spread of statistical education and training were unequal in the different developing countries, some being quite advanced and comparable to the developed countries and some at lower levels. Advantage of this heterogeneity could be taken by utilising the expertise available in those countries with higher levels of statistical education.

2. The gap between academic teaching of statistics in universities and colleges and the requirements of practitioners was universally noted. In some countries steps were being taken to achieve a closer collaboration between university departments and the official statistical system. Some participants were of the view that this was useful. Others suggested that universities might have two streams of curricula - academic and professional.

3. It was pointed out that the quality of data and statistical expertise depended on the degree of realisation by the governments of the usefulness of statistics for policy making. More and more countries were progressing towards this realisation.

4. Commendable work was being done in remote and inaccessible islands of the South Pacific.

5. School level education in statistics was either non-existent or in a rudimentary stage in most countries.

6. International bodies like the UNESCO should be approached for help in coordination, holding of meetings etc. and, in general, in helping in the spread of statistical education and training in developing countries.

TOPIC 11: CO-OPERATION BETWEEN ACADEMIC AND PRACTISING STATISTICIANS

ORGANISER: PROFESSOR W G HUNTER

CO-OPERATION BETWEEN UNIVERSITY AND INDUSTRY STATISTICIANS IN THE UNITED STATES

R D Snee, E I du Pont de Nemours and Company, Wilmington

INTRODUCTION

Rarely a week goes by without an article on cooperative interactions between academic and industrial communities appearing in Science, Nature, or other journals of science and engineering. This dialogue has been stimulated in large part by the decreasing growth of student populations and the reduced Federal funding for academic research. These developments have also been a concern to statisticians in the U.S. and the American Statistical Association (ASA) has formed committees to study the issues and to make recommendations (ASA 1980, ASA 1982). These reports have in turn produced comment by several university and industry statisticians (e.g. Bradley 1981, 1982; Hill 1982; Hunter 1981, 1982; Rustagi and Wolf 1982, Snee 1982).

Cooperation between industry and academia in the U.S. has a long history. There are many models to follow (e.g. Russell and Turner 1968; Wicks 1981) and new approaches continue to be proposed (Roy 1981). Tebo (1982) emphasizes that cooperative programs can be viewed as a mechanism of increasing university productivity. There has also been considerable cooperation in the field of statistics. It is apparent, however, that this activity must be increased and broadened if our profession is to successfully meet the challenges that face us. Certainly statistics has no God-given right to survive as a discipline and profession and we must continually look for ways to increase our impact and influence.

Various methods of interaction and cooperation are reviewed in this paper and suggestions for new approaches are presented for consideration. Other suggestions can be found in ASA (1980). It is apparent that the development of university-industry cooperative programs will result in a broader role for statisticians in our society. In Snee (1982) it was emphasized that focusing on opportunities such as training for industry, development of improved communication skills, statistical quality control training and research, and the establishment of consulting intern programs are all part of developing a broader view of statistics. Interactions with industry can help universities take full advantage of these opportunities.

KEYS TO SUCCESS

We must recognise at the very beginning that successful cooperative programs do not happen by themselves; they require careful planning and a recognition of some key ingredients that must be present (Table 1). Fruitful interactions begin with both parties getting to know each other and understanding each other's needs and objectives. It has been my experience that unsolicited proposals from unknown parties with no prior contact are almost always rejected out of hand. Industry must communicate who they are and what they do and suggest programs that are of interest to them. Universities must identify their goals and the constraints of their environment. I know of several instances in which both parties have held unrealistic views of the other's situation.

TABLE 1

COOPERATION BETWEEN UNIVERSITY AND INDUSTRY
CHARACTERISTICS OF SUCCESSFUL PROGRAMS

- Both parties must get to know each other.
- Either party can initiate the discussion.
- There must be mutual benefit to both parties.
- Interaction with nearby universities is easiest to implement.
- Programs can start small and include a plan for future growth.
- No single approach will work in all cases.

In order to get the interaction started someone must take the first step. Both parties have the responsibility to do this. If the proper groundwork has been done, it is not unlikely that the proposal will be met with the response, "we think that your proposal is worth pursuing; in fact, we have been thinking along the same lines ourselves. Let's get together to discuss it further". Experiences of other disciplines (e.g. Russell and Turner 1968; Tebo 1982) indicate that contacts and interactions between interested individuals at the university and company (frequently one person in each organization) are needed to define and carry out the cooperative programs. The success of the programs depends heavily on the dedication of the individuals involved.

It should also be recognized that both the university and the company must benefit from the interaction. Fusfeld (1980), Director, Center for Science and Technology Policy, New York University, in a Science editorial, points out that:

> "Initial approaches were made by universities with overtones of 'with your money and our brains' - not an endearing note, and surely not the best one on which to begin a relationship. ... Universities and industry are now building toward long-term relations that take into account each other's needs and contributions, the functions that each serves in our society. This is, in short, a sound base for a sensible working partnership".

E. G. Jefferson, Chairman of The Du Pont Company, in commenting on collaborative research efforts, emphasizes that these agreements

> "... are built on the sound foundation of mutual benefit, ..." and they emphasize "... not what industry can do for academia, but what we can do for each other".

There are many ways that industry can help beyond the awarding of unrestricted grants. Innovative thinking is needed to take full advantage of the possibilities.

The interaction is usually easiest to implement with nearby universities because there is greater opportunity for extended face-to-face contact, visits to the other party's site, adjunct faculty appointments, intern programs, etc. This greatly facilitates the "getting to know each other" process.

In getting a cooperative program off the ground it is usually easier to start small and include a plan for future growth. Starting small (e.g. one or two adjunct faculty appointments and a student intern program) provides flexibility and is easier for university administration and company management to approach. As these efforts prove successful and the parties get to know each other there will be stimulus and support for other programs such as joint research, visiting appointments, etc. A useful parallel is that of developing a consulting relationship with an organization. One usually begins with an important project with modest objectives. The successful completion of the effort often leads to other projects with broader scope and greater impact.

It is not possible to detail a prescription that will work in all situations. A variety of approaches is possible. Each program should be tailored to the needs of the organizations involved. The following sections describe various forms of cooperative programs that have been developed. The discussion is organized by those interactions that take place with a university and those outside the university including those through professional societies such as the ASA and the American Society for Quality Control (ASQC).

INTERACTIONS WITH A UNIVERSITY

Interactions between a university statistics department and a company can take a number of different forms (Table 2). One of the easiest programs to implement is visits to universities by industrial statisticians and vice versa.

77

Many universities and companies extend such invitations several times each year, usually for a day consisting of a seminar and one-on-one discussions. Visits of two days to one week are increasing in popularity and are particularly useful in getting a feel for what life is like in a given site. Visits for longer periods of time are also possible although they appear easier for university statisticians to make. Statisticians working in industry tend to have responsibilities that are not easily shifted from one person to another.

TABLE 2

INTERACTIONS WITH A UNIVERSITY

. Visiting lecturers from industry.
. Appointment of industry statisticians as adjunct faculty.
. Consulting intern programs in industry.
. Inclusion of industry statisticians on university advisory committees.
. Special programs and short courses.
. Problem solving for industry.

Industry statisticians often serve as adjunct faculty teaching courses in their specialty, helping out with teaching overloads, advising students, and directing thesis research. Students generally respond positively to the opportunity to take courses from statisticians who work on industrial problems on a regular basis and provide a different perspective on the role of the statistician.

Programs such as visiting lecturers from industry and appointment of industry statisticians as adjunct faculty members have been in existence at a few universities for many years. These programs will naturally expand as the interaction between universities and industries increases. Both parties should initiate discussions to speed up the process.

Employment of students in industry as consulting interns or as summer employees has been implemented by a few universities; however, it appears that much more can be done (e.g. Jones and Kanji 1980; Stoline 1982). It should be recognized that such programs require time to administer effectively and may be available to only a few students because an industrial organization can generally handle only one or two students per year. Consulting intern programs in industry may be difficult to implement if the university is far from an area with a large number of companies.

A university should also consider having a consulting intern program at the university in conjunction with its statistics laboratory (Snee 1982). Ideally, a university should have all three programs (consulting intern in industry or at the university and summer employee in industry) with students participating in those programs that will do them the most good. The key objective is for the student to work on real problems under the guidance of a statistician experienced in statistical practice (ASA 1980, Section 3.4). This can be done either in industry or at the university. I worked, during my graduate study, as a summer employee for Jones and Laughlin Steel Corporation and as a consulting intern at Rutgers University (Snee 1982). I found both of these experiences very worthwhile in learning the realities of statistical practice and how statistics is used in solving the problems of science and engineering.

Consulting intern programs of all types are of interest to industry and they enthusiastically endorse such activities. Through the efforts of William J. Hill, the Allied Foundation has provided funding for a Workshop on Consulting Intern Programs to be held at the University of Wisconsin in November 1982. The workshop, which was organized by Brian L. Joiner, includes leaders in consultant training from industry, government, and academia.

Some universities (e.g. Brigham Young, Ohio State, Oakland) are including industry statisticians on their departmental advisory and evaluation committees. This can be beneficial to both parties and should be pursued more actively. It can give the program added credibility with industry and be a source of ideas, energy, and publicity.

The activities discussed above are well known, if not widely implemented, and have received considerable discussion (e.g. ASA 1980). The opportunities associated with programs such as short courses for industry, research on unsolved statistical problems, and the investigation of subject matter problems are beginning to be recognized by universities and merit consideration and discussion by both universities and industry.

Special Programs and Short Courses for Industry

Statisticians working in industry have a need to keep up with recent developments in the field and to learn new technology. Brigham Young University recognized this need and has run a one-week Summer Institute on Applied Statistics since 1976. A well-known statistician serves as the primary lecturer with supplementary presentations by BYU faculty and statisticians working in the area. This institute benefits not only the persons who attend (mainly practising statisticians), many of which are BYU alumni, but also BYU faculty and students. Other universities should find this kind of program attractive because it gives the university's alumni an opportunity to return to the campus.

The University of Massachusetts offers each year a three-day Seminar on Statistics in Biopharmaceutical Research and a two-week Summer Program in Biostatistics. These programs are sponsored by the Biopharmaceutical Research Unit which is funded through grants from several pharmaceutical companies.

Scientists and engineers working in industry have a need to learn what statistics can do for them and their companies. The University of Wisconsin School of Engineering and Statistics Department held a two-day seminar on Statistics In Industry in February 1982 to fill this void and plans to present similar programs in the future. Speakers from Consolidated Papers, Du Pont, Kodak, 3M Industries, 3M Company, and the University of Wisconsin emphasized how statistics is used and the impact it can have in industry. Videotapes of the sessions are available for distribution. For further information, write Mr. Robert Perras, Kurt Wendt Engineering Library, University of Wisconsin, 215 N. Randall Avenue, Madison, WI 53706.

The University of Tennessee has focused on another critical industrial problem; that of improving product quality through the use of statistical quality control (ASA 1981). The University has formed The Institute for Productivity Through Quality which offers an intensive three-week training program spread over a three-month period (one week/month). Many more such programs are needed to satisfy the demand in the U.S. for training in statistical quality control. Universities can play a leadership role here.

Problem Solving for Industry

In developing cooperative programs with industry, it is logical to think of research projects on statistical methodology. Bradley (1981, 1982) suggests that industry should pose unsolved statistics problems for universities to investigate. This is certainly appropriate for, as Box (1980) points out, interest in practical problems has led many statisticians (e.g. Gauss, Gosset, Fisher, Wald) to make important theoretical advances. Companies are more likely to do so if they have been approached by a university expressing interest in such problems. Some persons argue that if a problem is important to a company, they will solve it themselves, perhaps with the aid of a consultant. A company is also unlikely to recognize a statistics research problem unless it has someone on its staff with a knowledge of statistics. It seems that it would be particularly worthwhile for a university to initiate contact with those companies whose personnel have little formal training in statistics.

The identification of important unsolved statistics problems can be done through our journals; however, a more effective method is for industry and academic statisticians to discuss problems of mutual interest in a conference/workshop setting. This could be done through a professional society or organized by a specific university or group of universities.

The Mathematics Department of Duke University is holding such a program that is patterned after a similar workshop held for more than ten years at

Oxford University in England (see SIAM News, June 1979). Problems are solicited from industry and government and reviewed by the faculty to determine interest in the problem and potential for solution. The proposers of the selected problems are invited to a week-long conference in which a small study group is assigned to tackle each problem. The conference meets as a group at the end of the week to discuss their progress. It is not unreasonable that the interaction and dialogue might continue after the conference.

The key ingredient in such an arrangement is university and industry statisticians who have a mutual interest. A small seminar might be a good way to continue the dialogue.

The solution of subject-matter problems is another opportunity for cooperative programs. Environmental studies is an area where there is tremendous need that can be filled by university statistics departments. Research on these problems can be funded through trade associations because the problems are of concern to several companies. These problems often require specialized statistical methodology and the development of new techniques. Such activity is characteristic of the university. Expertise and reputation in a particular area of statistics and an interest in working on practical problems are key considerations when industrial groups seek such help.

The University of Wisconsin did considerable work in the mid-1970's on the analysis of Los Angeles air pollution data with funding from the American Petroleum Institute (Tiao, Box, and Hamming 1975a, 1975b; Tiao, Phadke, and Box 1976). More recently, Wisconsin and Princeton University have made extensive analyses of trends in stratospheric ozone (Reinsel, et al 1981, Bloomfield, et al 1982) with support from the Chemical Manufacturers Association. Time series methodology played a key role in both of these projects.

INTERACTIONS THROUGH PROFESSIONAL SOCIETIES

Interaction between university and industry statisticians can also take place in conjunction with professional society activities (Table 3). These contacts enable both parties to get to know each other and provide a mechanism through which they can address concerns of mutual interest.

TABLE 3

INTERACTIONS THROUGH PROFESSIONAL SOCIETIES

- Industry support of professional society activities.
- Meetings with speakers from both industry and universities
- Conferences on unsolved problems and other key issues.
- Committees to address common problems.

Industry has a long history of support of professional and learned societies in the United States. Industry not only sends their statisticians to meetings but also supports their activities as meeting organisers, speakers, moderators, committee members, and society officers. Part of the ASA's program on filming of distinguished statisticians has been supported by the efforts and funding of the Pfizer, General Motors, and Du Pont Corporations. Industrial statisticians have also written numerous articles describing statistical consulting in industry (ASA 1980; Greenfield 1979; Hooke 1979; Marquardt 1979, 1981) and presenting results of their research and statistical studies (e.g. see *Biometrics*, *Technometrics*, *J. Quality Technology*).

Interaction and cooperation can be stimulated by insisting that conference committees and speaker lists include representatives from both academia and industry. This has long been a guiding principle of the Gordon Research Conference on Statistics in Chemistry and Chemical Engineering and the Fall Technical Conference of the Chemical Division of the American Society for Quality Control (ASQC) and the ASA Section on Physical and Engineering Sciences. A goal of the ASQC Chemical Division's short course program, which began in the 1950's, has been that the instructors in each course should be one person from industry and one person from a university. This ensured a variety of viewpoints and encouraged cooperation between university and industry. *Applied Regression Analysis*, the first text to address the practice of regression analysis, grew out of the notes for one of the ASQC Chemical Division's courses.

The key to success of any conference is a balanced program that has something for everyone. It is disappointing to see a meeting with an underrepresentation of one group or another. The balanced program is, of course, more difficult to develop and takes greater effort but is more valuable in the long run and is usually well attended.

Conferences also present unique opportunities for students to meet and get to know statisticians from industry. Universities can arrange for students to attend. It is usually possible to negotiate reduced registration fees or find ways for the students to help out in lieu of the fee. It was under such an arrangement that, as a graduate student, I first met R.W. Kennard of Du Pont who was to later hire me into Du Pont's Applied Statistics Group.

Two other examples of cooperative efforts through professional societies are the founding of *Technometrics*, a journal of statistics for the physical and engineering sciences, and the Biopharmaceutical Section of ASA. *Technometrics* was started with matching funds from the ASQC Chemical Division and ASA. The Biopharmaceutical Section of ASA resulted from the rapid growth of statistics in the pharmaceutical industry and the need for industry and university statisticians working on pharmaceutical problems to meet and exchange ideas and experiences.

In my visits to universities I have found that many faculty and students desire more information on statistics in industry. I have suggested that professional societies can help by developing a list of companies that employ statisticians, preparing a speakers list of statisticians in industry who are interested in visiting universities to discuss opportunities in industry, and by arranging workshops and question-and-answer panel discussions at local and national meetings concerning industry-academic programs and related issues (Snee 1980).

Professional societies can play an important role in getting university and industry statisticians together in conference and workshops to discuss key issues. A conference on unsolved statistical problems, as discussed in 'Problem Solving for Industry', could be sponsored by a professional society. Improved public perception concerning the role and contribution of statisticians and methods of recruiting students into the profession are possible topics for such conferences. These meetings could be held at either the local or national levels and could be in the form of "statistics days" in which students can learn about careers and opportunities in the field of statistics.

The formulation of committees of professional societies to address issues of common interest is another good way to develop cooperation between university and industry statisticians. The ASA Committee on Training of Statisticians for Industry (ASA 1980) is a good example where fruitful interactions have taken place. ASA has also recently formed a continuing education advisory committee composed of representatives of both industry and academia.

OTHER PROFESSIONS ARE FACING SIMILAR PROBLEMS

It was pointed out at the beginning that interest in cooperative programs between industry and universities is due in large part to decreasing student populations and reduced Federal funding for academic research. Statisticians must recognize that computer science and some fields of engineering are experiencing many of the same pressures as statistics and that companies will view statistics in competition with these fields when developing and funding cooperative programs. The Du Pont Company, for example, made a $6 million grant to Harvard University for research in genetic engineering (Stein 1981) and recently established a program to encourage chemical engineering students to pursue the PhD degree (Tebo 1982). Both of these areas are of keen interest to Du Pont and play a significant role in the future of the Company.

We must also keep in mind that statistics, particularly in industry, is a young profession (industrial statistics groups were first established in the U.S. in the early 1950's) relative to other fields of science and engineering. It will take many years (perhaps a generation or two) for statisticians to have the impact and gain the influence now enjoyed by other professions.

Statistical studies also result in a report rather than a piece of hardware (i.e., a machine or process). It has been my experience, and that of many others, that hardware always wins in the sense that it is easier to convince people of the utility of hardware and, in the case of computers, its associated software, than of concepts, theories, and procedures. Statisticians should take advantage of this by emphasizing the role of the computer both in implementing the procedures of statistical practice and in statistical research. Industry is a good source of examples of the use of the computer in statistical studies.

I make these comments not to discourage the profession but to point out the realities of the situation so that we are not disappointed in our progress because of unrealistic expectations. University-industry cooperation has benefited the profession in the past and increased activity in this area will serve to stimulate even greater advances in the future.

ACKNOWLEDGEMENTS

The author is indebted to D.W. Marquardt and P.V. Tebo for stimulating conversations regarding the interaction between universities and industry and for their helpful comments and suggestions on the presentation of this paper.

References

[1] ASA Committee on Training of Statisticians for Industry (1980) "Preparing Statisticians for Careers in Industry" (With Discussion). The American Statistician, 34, 65-80/

[2] ASA Committee on Training of Statisticians for Industry (1981) Letter to the Editor. The American Statistician, 35, 269. Need for training in statistical quality control is discussed.

[3] ASA Committee on Training of Statisticians for Federal Government (1982) "Preparing Statisticians for Careers in Federal Government". The American Statistician, 36, 69-89/

[4] Bloomfield, P., Thompson, M.L., Watson, G.S. and Zeger, S. (1982) "Frequency Domain Estimation of Trends in Stratospheric Ozone". J.Geophysical Research, In Press.

[5] Box, G.E.P. (1980) "Comment on Preparing of Statisticians for Careers in Industry" The American Statistician, 34, 76-77.

[6] Bradley, R.A. (1981) "What Industry Can Do For Statistics". Presented at the Fall Technical Conference on the ASQC Chemical Division and ASA Section on Physical and Engineering Sciences, Gatlinburg, TN, October 1981.

[7] Bradley, R.A. (1982) "The Future of Statistics as a Discipline" J.American Statistical Association, 77, 1-10.

[8] Draper, N.R. and Smith, H. Jr. (1966) *Applied Regression Analysis*, John Wiley and Sons, Inc., New York. Second Edition appeared in 1981.

[9] Fusfeld, H.I. (1980) "The Bridge Between University and Industry". *Science*, 209, No.4453, 221.

[10] Greenfield, A.A. (1979) "Statisticians In Industrial Research: The Role and Training of the Industrial Consultant". *The Statistician*, 28, 71-82.

[11] Hill, W.J. (1982) "Training of Statisticians for Employment in Industry". In *Proceedings of the Ohio State University Conference on Teaching of Statistics and Statistical Consulting*, eds. J.S. Rustagi and D.A. Wolfe, New York: Academic Press.

[12] Hooke, R. (1979) "Getting People to Use Statistics as God and Sir Ronald Fisher Intended". *The American Statistician*, 33, 39-42.

[13] Hunter, W.G. (1981) "The Practice of Statistics: The Real World is an Idea Whose Time Has Come". *The American Statistician*, 35, 71-76.

[14] Hunter, W.G. (1982) "Graduate Programs in Statistics". In *Proceedings of the Ohio State University Conference on Teaching of Statistics and Statistical Consulting*, eds. J.S. Rustagi and D.A. Wolfe, New York: Academic Press.

[15] Jefferson, E.G. (1982) "Chemical Research: An Industry View". *CHEMTECH*, 12, 260-262.

[16] Jones, J.D. and Kanji, G.K. (1980) "The Role of Professional Experience in Statistical Education". *The Statistician*, 29, 196-203.

[17] Marquardt, D.W. (1979) "Statistical Consulting In Industry". *The American Statistician*, 33, 102-107.

[18] Marquardt, D.W. (1981) "Criteria for Evaluating Statistical Consultants in Industry". *The American Statistician*, 35, 216-219.

[19] Reinsel, G., Tiao, G.C., Wang, M.N., Lewis, R. and Nychka, D. (1981) "Statistical Analysis of Stratospheric Ozone Data for the Detection of Trend". *Atmospheric Environment*, 15, 1569-1577.

[20] Roy, R. (1981) "Graduate Universities - A New Model". *Science*, 214, 1177.

[21] Russell, T.W.F. and Turner, H.E. (1968) "Internship in Chemical Engineering Design". *Chemical Engineering Education*, Spring 1977, 74-77.

[22] Rustagi, J.S. and Wolfe, D.A. (1982) *Proceedings of the Ohio State University Conference on Teaching of Statistics and Statistical Consulting*, New York: Academic Press

[23] Snee, R.D. (1980) "The Bridge Between University and Industry". *Amstat News*, No.70, December, 1.

[24] Snee, R.D. (1982) "Comment on Preparing Statisticians for Careers in Federal Government". *The American Statistician*, 36, 82-87. Comparison of training needed by industry and government is made.

[25] Stein, M.D. (1981) "Industry and Research: Du Pont Joins In". *Nature*, 292, 191.

[26] Stoline, M.R. (1982) Western Michigan University Biostatistics Internship Program. Personal Communication.

[27] Tebo, P.V. (1982) "Programs Contributing to Increased University Productivity: A Grass Roots Approach". Presented at The American Society for Engineering Education, Texas A&M University, College Station, TX, June 23, 1982. The Du Pont Company's university financial assistance and partnership programs are discussed.

[28] Tiao, G.C., Box, G.E.P. and Hamming, W.J. (1975a) "Analysis of the Los Angeles Photochemical Smog Data: A Statistical Overview". J. Air Pollution Control Association, 25, 260-268.

[29] Tiao, G.C., Box, G.E.P. and Hamming, W.J. (1975b) "A Statistical Analysis of the Los Angeles Ambient Carbon Monoxide Data 1955-1972". J. Air Pollution Control Association, 25, 1129-1136.

[30] Tiao, G.C., Phadke, M.S. and Box, G.E.P. (1976) "Some Empirical Models for the Los Angeles Photochemical Smog Data". J. Air Pollution Control Assocaition, 26, 485-490.

[31] Wicks, Z.W. Jr. (1981) "Industry and Academia Can Work Together - A Case History. CHEMTECH, 11, 740-745.

SOME OBSERVATIONS FROM EUROPE

Lars-Erik Öller, Ministry of Finance, Helsinki

STATISTICS AND SEMANTICS

To the best of my knowledge, in all Indo-European languages the word "statistics" is used to denote two different things, namely a collection of data *and* an academic discipline. In the only non-Indo-European language I know, i.e. in Finnish, there are separate words for the two concepts. (*)

I think this semantic flaw has resulted in a lot of confusion. There is, and should be, a difference between the aims and means of a statistical office and those of a research oriented statistical unit. The former has to meet the demand of society for certain data. The main task of the office is to collect those data and make them available to the public. This involves application of some statistical techniques, such as sampling theory, but to a much larger extent it requires subject matter insight and thereby other disciplines. National accounts imply knowledge of economics, social statistics is supported by sociology, meteorological statistics needs meteorology etc.

If statisticians think that they and only they are fit to assume full responsibility for routine processing of social data I think we are aiming at the wrong target.

Instead of getting involved in semantics I suggest we concentrate our efforts on three areas where the need for statistical knowhow, if not overt, at least potentially, seems to be most urgent. These are: *digestion of existing data*, *general assistance in empirical research* and *developing better methods to achieve these goals*.

DIGESTING EXISTING DATA

In an industrialized European country, like the one I represent, a vast amount of data is produced by the central statistical office. It is very hard to give a precise figure of how much data is supplied. A moderate estimate is that two million time series are published in print each year. On request 20-30 times that amount is available to the public. The office has been working for nearly one hundred years.

These figures apply more to the level of development of a country than to its size. The latter affects more the magnitudes of the figures than the amount of data.

(*) "Tilasto" meaning a collection of data and "tilastotiede" meaning the science for (a collection of) data.

Since further analysis of this data mainly takes place outside the production unit it is even harder to say how much the data is further analyzed. My guess is that an extremely small fraction of it is ever analyzed using statistical methods more sophisticated than comparing the magnitudes of a few figures. Averages are sometimes applied but measures of disperson are almost unknown. In some cases graphical displays are used. Ad hoc time series methods, such as seasonal adjustment, are common in processing economic time series. The amount of time series that would go into any more advanced statistical processing probably is closer to a three digit number than to a five digit one.

The core of statistical theory is, in my view, making inference from quantitative observations. In parametric statistics we are interested in distributional properties of data, such as location, uni- and multivariate dispersion, skewness etc. Available data provides limitless possibilities to apply what methods we know and to develop new ones.

To give a concrete example of the possibilities let us think of a manageable collection of time series. Their joint distribution can be classified into three main cases:

(i) Each series is white noise and they are mutually independent for all lags,

(ii) The series are mutually independent but they are serially correlated,

(iii) The time series are both serially and cross correlated.

I will study these cases from a time series analyst's point of view. The best a statistician can do in case (i) is to show that the variables can be regarded as white noise, show what the series look like and estimate the means and the variances.

In case (ii) a statistician can apply univariate ARIMA techniques to find a model for each time series. The models act as filters that turn mutually independent autocorrelated time series into independent white noise series, i.e. to case (i). The models can be used for forecasting. ARIMA and spectral techniques can be combined for decomposing the series (seasonal adjustment) etc.

When joint dependence prevails between the time series, as in case (iii) they can be filtered into mutually independent white noise using e.g. multivariate ARIMA techniques. Thereby we may get an idea about the true behavioural mechanisms that are reflected in the data and we would be able to simulate them.

It is rather difficult to find good examples of time series that are to be regarded as mutually independent white noise. However, just estimating a mean and maybe a variance means that from a statistician's point of view we are acting as if the data were white noise.

It took us a century to build the data base we have today. Let's now start building a research structure on this base. Processing equipment is becoming better and cheaper every day and we possess the necessary knowhow to start building. An easy starting point is to build univariate ARIMA models producing forecasts that would routinely be published together with observed data. A more distant goal is multivariate models, the errors (shocks) of which may be of enough interest to be published together with the observed series. Decision makers would focus their interest on impulses, not directly osbervable, reaching the system and on structural changes more than on the normal, model-specified working of society.

ASSISTANCE IN EMPIRICAL RESEARCH

It is hard to think of any empirical research that could not benefit from statistics. In fact scientific experiment is the cradle of modern statistical thought as introduced by R.A. Fisher. Still, I have the impression that in many, if not in most, empirical research teams the professional statistician is absent. We may ask: how do biologists, physicists, economists etc. manage to do empirical research without the assistance of statisticians?

As I see it there are three ways to avoid a statistician:

(i) You conduct your experiment as much as possible in the same way as others have done before you. Assuming that the method is applicable in your case the only thing that can go wrong is that the method was wrong in the first place. But then you would be in good company.

(ii) Empirical scholars are given basic training in statistics. A biologist can apply a t-test, a physicist knows something about stochastic processes and an economist has taken a basic course in econometrics. Instead of using this knowledge as a working language for communicating with a statistician you can stick to the rule of "do it yourself".

(iii) Collect your data as well as you can and bring it to the computer center. They will have somebody with some statistical training who knows all the ready-made program packages. At least one will fit your purposes.

Why would people avoid statisticians? I see two important reasons: *ignorance* and *convenience*. Many statisticians have a large share in the first sin and we are not completely innocent on the second account either.

Ignorance first. People simply do not know that we could be of assistance. Too few statisticians have endeavoured to market their knowhow. For that it is necessary to meet the empiricist halfway by learning at least as much about his/her subject as he/she knows about statistics.

Convenience. It is more convenient for the empiricist to try some of the avoiding tricks than to have to rely on a strange theoretician who may even turn all you have done upside down and with whom you would have to share the glory. For the statistician it may be much more convenient to indulge in some theoretical problem that few, if any, people can understand and criticize. This kind of research gets green light more because of its contribution to other statisticians' feeling of ignorance than due to its contribution to knowledge.

What can we do to improve the situation? I think that the answer lies in changing attitudes, ours and theirs. Every statistician should get training in soiling his hands with real - not only textbook - data. In practice this implies that every statistics department practices consulting where students under the supervision of senior statisticians, help to solve empirical research problems. The students would also have to acquire at least basic knowledge of an applied field. On the other hand, students of an applied field would get accustomed to the idea of conducting joint research with statisticians. This policy aims at changing attitudes where they can be changed: in young age groups.

METHODS

If some of the criticisms of theoretically minded statisticians for being ignorant and arrogant can be well founded, some of it may be beside the point. No theories are final and methods applied in practice are often just the tip of an iceberg of theory. During the development of that theory it was not always clear where it could be put into practice.

Practitioners must realize that there is no substitute available for mathematical language. It is an auxiliary device that helps us to understand. Those for whom mathematics is an end in itself would more properly be called mathematicians and those at the other end who substitute numerical observations for symbolic variables are applied people in the subject proper. Both are needed but let's keep the road open between the two ends.

These thoughts could be summarized in one of our favorite gadgets, the two by two table:

World Theory	simple	complicated
simple	A	B
complicated	C	D

Sometimes our problems are simple and require simple solutions (A) but maybe more often the world is complicated and problem solution requires complicated theory (D). Suboptimal behaviour occurs in the off diagonal entries. In C we are talking beside the point and in B we want to see the world as simpler than it really is. It is not always just theory that is hard to understand - it may be the world.

DISCUSSION

The theme of this meeting is a popular one. In Scandinavia alone it has been a topic of discussion at four international meetings since 1970 (cf. Statistical Review of the Swedish National Central Bureau of Statistics 2/1978). On a world scale the discussion has been going on for decades. Many of the ideas put forward at this meeting have probably been expressed before. I have no idea about the effect these discussions may have had. The time has come to talk about the implementation of all the good suggestions.

Here the International Statistical Institute (I.S.I.) comes to my mind. It has the necessary authority to issue international recommendations that could improve and coordinate the teaching of statistics.

The I.S.I. has traditionally published bibliographical studies. It could, as a next step, help to improve the status of statistics by publishing papers on the history of statistics. Not only should every statistician know the historical background of his knowledge but in a popular form this narrative should be distributed as widely as possible. It should be as natural to connect Bayes and Fisher (*) with statistics as Linneaeus with botany and Keynes with economics.

Finally, the discussions about the role of statistics have a social aspect. This is the occasion when we can laugh at ourselves. We go as far as quoting classical poems, such as those about slaying dragons and identifying elephants. In the following poem I have tried to contribute to that tradition:

(*) Cf. J. Fisher-Box: "R.A. Fisher, the Life of a Scientist", John Wiley and Sons Inc., New York, 1978.

STATISTICS

Statistics to the uninformed
Is columns, rows that on us stormed
When handling data became the feat
Of dilettantes you couldn't beat
In asking questions about you
When you had many things to do.

But the informed would simply frown
And tell a story, much of his own
Of what statistics really is
(Letting all his s's fizz).
And not without a bit of pride
He thus would unto you confide:

No! Statisticians never bother you.
They write no numbers, except maybe one or two
Expressing confidence in what they believe
(Some do not even these conceive).
They assume the world has been or is to be observed
By which their peace of mind can greatly be preserved.

They use their hieroglyphs to show where you should look
(Rules for the design can be found in many a big book).
And when the observations have been picked
They tell you how to make inference, how to predict,
And suddenly a chaos of complexity
Seems to have turned into a system of simplicity.

Alas! The world does not always fit into our abstraction
And this can hardly be improved by mere obstruction.
So let's not forget the guy messed up in data.
And if, for fun, we sometimes get absorbed by *epsilon* and *eta*
At least it should be something new and true and nice
For such extravagance to be worth its price.

BUILDING BRIDGES BETWEEN THE ACADEMIC AND REAL WORLDS - SOME OBSERVATIONS FROM SOUTH AMERICA

C A Peres and P A Morettin, University of São Paulo

INTRODUCTION

The purpose of this paper is to summarize the present state of the statistical education and profession in South America, with special emphasis on Brazil. We analyze the problems of statistical education, training and research and indicate who the main users of the statistical methodology are.

In recent years some attention has been devoted to the problem of closing the existing gap between the academic statisticians and the professional statisticians working in government agencies, industries, etc. We refer in particular to the Conference on Transfer of Methodology Between Academic and Government Statisticians (1978), sponsored by the American Statistical Association (which will be referred to briefly as CTM).

Concerned especially with Latin America we mention the SELENES (Latin-American Seminar on the Teaching of Statistics) (1978) and a Meeting of the Working Group to Study the Status of the Statistical Profession in Western Hemisphere (1980), supported by the Inter-American Statistical Institute (IASI) (referred as WG for brevity).

We conclude by discussing an experience at the University of São Paulo, showing how a consulting course, which is taught to undergraduate students, can help as one means of closing this gap betwen the academic and real worlds.

EDUCATION AND TRAINING OF STATISTICAL PERSONNEL

In this section we try to summarize some sparse information, available from different sources, on the stage of the development of the education and training of statisticians in South America.

By education we mean that an academic degree is granted, at the undergraduate or graduate level. By training it is meant that special courses are offered in order to supply limited knowledge for special tasks (see Gura, 1980). Usually, this training is an in-service training, that is, given after the employee is employed by the organization. See CTM, Chapter IV.

Murthy (1978) states that the objectives of education are "to develop a statistical sense with a view to understanding socio-economic situations and natural and physical phenomena through statistics and teaching the methodology appropriate to produce statistics economically and to analyse and use them meaningfully", while the objective of training is "to impart special skills required to plan and undertake statistical operations necessary in producing and analyzing statistics".

Concerning undergraduate education in South America, thirteen countries have programs in Statistics; in Brazil, seventeen schools have such programs. At the graduate level, seven countries have programs leading to a master's degree and none to a doctoral degree. In the case of Brazil, seven centers offer graduate study at master's level. See Appendix A and B.

Two main conclusions were drawn in WG referring to education and training:

(a) "The principal limiting factor for the establishment of appropriate programs is the absence of teaching personnel adequately qualified for such responsibilities. As a result, the students participating in various programs, frequently receive low quality training or training that is not oriented toward the needs of government and industry, the most likely sources of employment;

(b) the absence of well qualified teaching personnel is principally due to the scarcity of resources allocated to graduate statistical training in most of the countries of the region".

In some countries (Brazil and Chile, for example), efforts are being made to implement reciprocal cooperation with more advanced centers in Europe and North America, with the objective of establishing doctoral programs.

Due to this lack of graduate programs (at least at the doctoral level), the solution has been to send people abroad to obtain advance training. This is of course costly and many countries do not have the resources to support adequately such programs. This is also a long run policy and as mentioned by Bradley (1982a) "the first priority for the use of available qualified statisticians should be the well-supported staffing of needed training programs in Statistics, programs developed centrally on a national basis or, when appropriate or necessary, on a regional basis".

Statistics is relatively recent as a discipline. Only in the last three decades university education has had an increasing expansion. Due to historical factors, the teaching of statistical disciplines were done in Mathematics departments, hence the principal emphasis has been on the theoretical side. The implication is that the students receiving this biased formation do not fill the needs of industry or government offices.

Most of the countries try to provide "ad hoc" training facilities, generally through National Statistical Offices, in order to prepare the personnel for their needs. The exception, perhaps unique in South America, is a course offered by CIENES, called "Program in Statistical Techniques", which is intended to train those persons who will work in statistical offices in government.

In some countries the needs vary from one region to another. In any event, as we mentioned above, the government seems to be the main source of jobs for statisticians coming out of universities.

In this connection it is very important that the curriculum of statistical programs at the universities should be composed of disciplines relevant to the immediate applications of the region. For example, if the principal needs are statisticians for statistical offices, a curriculum centered in economics seems to be the adequate answer, with courses including principles of sampling and survey design, survey methods, construction of tables, graphical methods, data analysis, etc. (CTM, Chapter III). If industry is a major source of jobs, the emphasis should be on regression analysis, design of experiments, response surfaces, quality control, time series, etc. (Marquardt, 1979).

A way to balance the statistical education of future statisticians is suggested in Appendix C. The idea is to use a flexible structure of modules in order that the students acquire a substantial subject-matter content from other fields, as social sciences, economics, biology, etc. In one semester the students receive the basic information in the field (Economics, Epidemiology, Psychology etc.) and in the following semester a statistically oriented course (Econometrics, Biometrics, Psychometrics, etc., respectively).

Bradley (1982a) advocates that "on-campus statistical consulting in support of university research programs may be initiated". This recommendation is further discussed in "A Suggestion for Building the Bridge between the Academic and Real Worlds" further on.

A point of apprehension is that the enrolments in statistical courses is decreasing in the last years. This is discussed by Bradley (1982b) for the USA case, but it is true for South American countries also. A recent survey among freshmen in Mathematical Sciences at the University of Sao Paulo, showed that 80% of the students prefer to pursue studies in Computer Sciences, while only 6% indicated a preference for Statistics.

This fact is part of an overall situation, where enrolment of talented students in Mathematical Science courses (except Computer Science) has been decreasing steadily, due mainly to the low salaries paid to high school and university professionals, compared with those paid by industries, for example.

USERS OF STATISTICS

The major sources of jobs for statisticians in South America are the government and the universities. This is true even in heavy industrialized regions, as São Paulo, Brazil, where industry, commerce and market research could generate large amounts of jobs.

The main reason for this situation is the lack of knowledge of most other sectors that a statistician could be very useful in helping them increase productivity and maximize profit. They prefer to hire an economist or an engineer, to perform a job which is better suited to a statistician.

We think there is not enough promotion of the statistical profession in South American countries to show the importance of the statistician in any decision-making process in industry, government, management, etc.

A further drawback is that most of the industries that operate in the region are multinational companies, and they bring from their home countries all the necessary technology to function, including the statistical technology. This introduces a further limitation in the work of statisticians and surprisingly enough, in some instances, the supply of statisticians is greater than the demand.

Of course the topic on promotion of the statistical profession is intimately tied to other factors already mentioned, like education and training, the existence of professional and scientific meetings on the field, statistical publications, and others.

From what has been said before, it might be argued that an effective promotion of the statistical profession could be counter-productive, since on the whole we would not have, in the South American region, enough statisticians prepared to reach the demand for them. On the other hand, in some areas, the number of statisticians coming out from undergraduate courses each year is more than enough to meet the demand.

To further confuse the situation, in Brazil Statistics is a profession regulated by law, as are engineering and medicine, and only those statisticians who had their B.S. in Statistics have permission to work professionally, the only exception being the university personnel. It follows that most of the professionals working in industry, government offices and other places besides universities have only a B.S. in Statistics, while those who pursued some graduate program usually work in universities.

It is a fact that some universities in the USA, Europe and other countries are trying to reformulate their master's programs in Statistics, in order to prepare the graduate students to attack real world problems. That means that the master's program usually has a more "applied" character. Even if these students did not have an undergraduate degree in Statistics, they will have a greater maturity than those students who had an undergraduate program in Statistics and immediately after coming out of the university, find a job in industry or government.

In this connection, to try to remedy the situation caused by this regulation, an experiment was introduced in the Statistics undergraduate program at the University of São Paulo, which we discuss in the next section.

A SUGGESTION FOR BUILDING THE BRIDGE BETWEEN THE ACADEMIC AND REAL WORLDS

The undergraduate program leading to a B.S. degree in Statistics, at the University of São Paulo, is a four-year program, the first year being a basic one for all mathematical science students (see Appendix C).

A two-semester course in consulting was introduced in 1978, in the last year of the program. We describe now how this course was designed and make some comments on its results.

Two members of the staff of the Statistics Department are responsible for the course, each year, although every professor can supervise the statistical analysis of some project. The course is composed of three parts.

Part I - Design and analysis of research projects developed in several areas of the University;

Part II - Training period in some firm or government office;

Part III - Development of statistical methodologies used frequently in Part I and II and not included in the undergraduate curriculum (complementary subject).

The Statistics Department has a Statistical Laboratory, where Part I takes place.

Description of Part I

i) A researcher (from now on called simply a "user") comes to the Statistical Laboratory and fills in appropriate forms with information about himself, the objective of the research and the research itself;

ii) each project is supervised by one member of the staff of the Statistics Department and performed by two students. An appropriate statistical methodology is chosen, after an interview with the user;

iii) the supervisor and the students make an outline of the project to all students of the consulting course. The user and the professors responsible for the course attend the exposition. During the discussion the students usually ask many questions and the user plays an important role in the discussion;

iv) after the statistical analysis is completed, the students have to write a report (always in close contact with the supervisor), which is submitted to the appreciation of the user before it is typed;

v) a copy of the original data and a copy of the report (in the form of an "Applied Statistics Report") are kept in the files of the Statistical Laboratory. Usually they can be used as "case studies" in classes of several disciplines.

The report does not have a fixed structure, in order to motivate the creativity of the students. But usually they contain the description of the experiment, the statistical analysis, a summary of the conclusions and a reference list. In Appendix D we present a table listing the projects concluded in 1981.

Description of Part II

Since Statistics is a professional career in Brazil, the undergraduate program must contain a minimum of disciplines regulated by federal law. This includes a training period during two semesters. Most of the students do this training period in government offices, private banks and a few in industries. At the end of each semester we send questionnaires to these places with the purpose of evaluating the performance of the students.

Description of Part III

Part of the work load of the consulting course (2 hours) is devoted to lectures covering statistical methodologies not covered during the normal academic program. For example, this year, the lectures include topics such as categorical data analysis and cluster analysis.

We remark that the time spent in the training period is extra to the class period. The final grade of the student in each course is an average of those in all three activities discussed above.

FINAL COMMENTS

It is a fact (WG, p.5) that statistical activities in South America are basically carried out at the academic and government level, and to close the gap between the objectives of each one, it is recommended that a greater quantity of subject matter useful in the governmental area should be included in the academic curricula. Moreover, in those areas where industry, banking, commerce, market research may be sources of jobs, a balanced curriculum should be taught.

It is fundamental that the statistical profession be promoted adequately, and one appropriate tool for this purpose is the establishment of national statistical societies, holding regular scientific meetings and providing a substantial coverage of all sectors of the profession. Publications at the national or regional level are also of great importance.

The curricula of the university level statistical training should offer an appropriate balance of theory and methodology, together with practical work, carried out either in agencies that generate official statistics or in other sectors of society that use statistical techniques.

A suggestion is made of a consulting course which has shown to be of great value in order to achieve the goals mentioned above.

References

[1] American Statistical Association (1978) Report on the ASA Conference on Transfer of Methodology Between Academic and Governmental Statisticians. Washington, D.C.

[2] Bradley, R.A. (1982a) Statistical Training in Developing Countries, Int. Statist. Educ. Newsletter, Vol.2, No.2, a publication of ISI.

[3] Bradley, R.A. (1982b) The Future of Statistics as a Discipline, Journal of the American Statistical Association, Vol.77, p.1.

[4] Gura, B. (1980) La Profession Estadistica en el Hemisférío Occidental: Situación, Educación y Capacitación (Con Referencia Especial a la America Latina), IASI, Washington, D.C.

[5] Interamerican Statistical Institute (1980) Report on the Meeting of the Working Group to Study the Status of the Statistical Profession in the Western Hemisphere, Santiago, Chile.

[6] Marquardt, D.W. (1979) Statistical Consulting in Industry, The American Statistician, Vol.33, p.102.

[7] Murthy, M.N. (1978) Statistical Education, Training and Research and Their Utilization, in CTM, Chapter III, p.68.

[8] Seminário Latinoamericano Sobre Enseñanza de Estadistica (1978) Informe Final, CIENES, Santiago, Chile.

APPENDIX A

Schools with a Statistical Program in Brazil, Undergraduate Level (by state)

1. São Paulo - University of São Paulo
   Federal University of São Carlos
   State University of Campinas
   Faculdade Paes de Barros

2. Rio de Janeiro - National School of Statistics
   Federal University of Rio de Janeiro
   State University of Rio de Janeiro
   Catholic University at Rio de Janeiro

3. Pernambuco - Federal University of Pernambuco
   Catholic University at Recife

4. Ceará - Federal University of Ceará

5. Rio Grande do Norte - Federal University of Rio Grande do Norte

6. Minas Gerais - Federal University of Minas Gerais

7. Paraná - Federal University of Paraná

8. Bahia - School of Statistics of Bahia

9. Rio Grande do Sul - Federal University of Rio Grande do Sul

10. Brasília - University of Brasília

11. Paraíba - Federal University of Paraíba

APPENDIX B

Schools With a Master's Program in Statistics (by state)

1. São Paulo - University of São Paulo
   University of São Paulo at São Carlos
   University of São Paulo at Ribeirão Preto (Biostatistics)
   University of São Paulo at Piracicaba (Agriculture)
   State University of Campinas
2. Rio de Janeiro - Federal University of Rio de Janeiro
   Institute of Pure and Applied Mathematics,
   IMPA (National Research Council)
3. Pernambuco - Federal University of Pernambuco
4. Brasília - University of Brasília

APPENDIX C

Suggestion of a Curriculum for an Undergraduate Course in Statistics
(Where the main demands are for statisticians working in industry and government)

- Basic for all students in Mathematical Sciences

1st SEMESTER - Calculus I
Vector and Geometry
Introduction to Computer Science I
Introduction to Statistics I
Physics I

2nd SEMESTER - Calculus II
Linear Algebra I
Introduction to Statistics II
Introduction to Computer Science II
Physics II

- B.S. in Statistics

3rd SEMESTER - Calculus III
Linear Algebra II
Computer Programming
Statistical Inference I

4th SEMESTER - Calculus IV
Complex Variables
Exploratory Data Analysis
Probability
Statistical Inference II

5th SEMESTER - Introduction to Stochastic Processes
Nonparametrics
Sampling Techniques
Design of Experiments I

6th SEMESTER - Design of Experiments II
Regression Analysis
Multivariate Analysis
Applied Stochastic Proesses

7th SEMESTER - Statistical Consulting I
Modulus I
Modulus II
Modulus III

8th SEMESTER - Time Series Analysis
Statistical Consulting II
Modulus I
Modulus II
Modulus III

The disciplines in the modules may vary from one year to another. An example follows:

Module I (subject-matter discipline allied to a discipline with a statistical character)

Microeconomics and Econometrics
Epidemiology and Biometry
Vital Statistics and Demography
Psychology and Psychometrics
Biology and Biometry

Module II (Quantitative specialization)

(Math.)	Mathematical Analysis I and Mathematical Analysis II
(Bus. & Econ.)	Operations Research I and Operations Research II Accounting I and Accounting II
(Eng.)	Industrial Organization I and Industrial Organization II
(Comp. Sci.)	Mathematical Programming and Simulation Mathematical Programming and Graph Theory
(Physics)	Statistical Mechanics and Statistical Methods in Experimental Physics

Module III (Disciplines in the Statistics Area)

Game Theory, Information Theory, Decision Theory, Statistics in Market Research, Categorical Data Analysis, etc.

APPENDIX D

Research Projects, 1981 - Statistical Laboratory, University of São Paulo

PURPOSE AREA	Research Leading to Doctor Degree	 Master Degree	 Publication	 TOTAL
Architecture	-	-	1	1
Marine Biology	-	-	1	1
Biosciences	2	3	1	6
Botany	1	-	-	1
Biomedical Sciences	2	-	1	3
Pharmaceutical Sc.	2	2	-	4
Education	-	-	1	1
Nursing	-	2	1	3
Engineering	-	-	1	1
Social Sciences	1	-	1	2
Medicine	4	3	4	11
Dentistry	-	2	1	3
Psychology	1	5	1	7
Public Health	1	-	-	1
TOTAL	14	17	14	45

BUILDING BRIDGES BETWEEN THE ACADEMIC AND REAL WORLDS

B Jones, B J T Morgan and G B Wetherill, University of Kent at Canterbury

INTRODUCTION

In the following we describe how statisticians at the University of Kent have, over the last ten years, managed to bridge the gap between the 'academic' and 'real' worlds. This has not been accomplished easily but by dint of very hard work. Our description is divided into five sections. In the first of these we list some of the collaborative Ph.D. projects that have been undertaken by students working in the Department and in the following two sections we discuss some of the advantages and disadvantages associated with this type of collaboration. Following this we mention other types of collaboration and end with some final remarks. A list of references to some of the published work done by the staff and students at Kent is given at the end.

TITLES OF SOME Ph.D. PROJECTS UNDERTAKEN WITH OUTSIDE BODIES

1. S.H. Bishop - Three-way and higher cross-over designs in clinical trials - Pfizer Central Research, Sandwich, Kent.
2. P. Bravo - Acceptance sampling plans by variables for percent defective, with emphasis on double sampling plans - in association with the Ministry of Defence.
3. J.B. Curram - Modelling case studies in economic quality control - Amersham International Ltd., Buckinghamshire.
4. C. Daffin - Statistical studies in long term economic forecasting - I.C.I. Mond Division, Runcorn, Cheshire.
5. K.V. Darby - Statistical examination of British Bird observatories data - B.T.O., Tring, Hertfordshire.
6. P. Duncombe - Some problems in writing user-friendly intelligent regression software - I.C.I. Mond Division, Runcorn, Cheshire.
7. T.M.M. Farley - Statistical problems of chronic toxicity - Pfizer Central Research, Sandwich, Kent.
8. J.P. Hinde - Statistical principles of quality control procedures for saturation analysis assay kits - Amersham International Ltd., Buckinghamshire.
9. B. Jones - computer assisted design of experiments - I.C.I., Mond Division, Runcorn, Cheshire.
10. P.M. North - Statistical methods in ornithology - B.T.O., Tring, Hertfordshire.
11. C.C. Robertson - Statistical applications in memory studies - In association with a number of psychologists at Cambridge and Oxford.

12. S.A. Watts - Stochastic aspects in the use of compartment models related to pesticide evaluation - Shell Research Ltd., Sittingbourne, Kent.

THE ADVANTAGES

Statistics is a practical subject and through collaborative work we have been trying to develop practical statistical methods. For instance the two research studentships conducted in conjunction with the British Trust for Ornithology involved large sets of data which had been gathered by the Trust but not looked at in any detail and certainly not by a statistician. Results of the collaboration here have been a variety of new approaches, new methods and research papers, all of which have helped to bring the data alive. In this case, a large number of voluntary workers are involved in the data collection and it is extremely important that they should receive some form of feedback on the interpretation of the data. Birds are important indicators of changes in the environment and statistical methods here are clearly playing a wider, more important, rôle. Out of this collaboration arose a recent One-Day Biometrics Conference in London on "Statistical Methods in Ornithology". The research project of Simon Watts, in collaboration with Shell Research at Sittingbourne, dealt with stochastic aspects in the use of compartment models for describing experiments to investigate pesticide evaluation. Here again Statistics was being used with the protection and monitoring of the environment in mind.

The results of a collaboration can also be the development of useful tools which non-academic statisticians can use as part of their day to day work. For example, algorithms to search for optimal experimental designs were developed as part of a project with I.C.I. These algorithms are being used in a number of centres worldwide. The Caribbean Agricultural Research and Development Institute in Trinidad have versions which run on quite small computers and are used regularly to find designs for non-standard situations. A further example is provided by another I.C.I. project in which a user-friendly, interactive computer program has been developed for analysing regression data. The main feature of the program is that it has been developed for use by non-statisticians. In the development of this program the student worked as part of a larger team but nevertheless played a major rôle.

The contact with I.C.I. for example has led to further research contracts to extend the work done by the students. This extended contract has led to a useful working relationship which has proved profitable to both sides. As is apparent from the list in section 2, a number of contacts have supported more than one project over the years.

THE DISADVANTAGES

It is difficult to set up collaborative awards. A large amount of groundwork has to be undertaken and goodwill has to be established and maintained with the collaborating body. This frequently involves the production of regular research projects, which, in the last resort, do help the student with the writing-up stage. Furthermore a large amount of travelling may be involved, if, for instance, the collaborating body is in Runcorn or Oxford or in Tring, say. Another disadvantage is that it may not always be possible to follow-up the research and the result of a CASE Award may just be the scratching of the surface, the opening up of areas which may then not be developed to their full potential. A final disadvantage is that these projects, when the aim is for a Ph.D., do have to try to satisfy a double aim, that is, the student wants to produce an academically respectable and acceptable dissertation but, at the same time, the collaborative body has to be kept happy. This suggests that the period of study should, in fact, be longer than the usual three years and a further pointer to the same conclusion lies in the need of the student to spend time and regular visits with the collaborating body, as well as spend quite possibly an appreciable amount of time becoming familiar with the practical background, quite apart from the necessary study of the theoretical background.

Supervision of these projects can also be additionally difficult and time-consuming since the supervisor also has to become acquainted with what could be an unfamiliar area.

There are also other disadvantages, not the least of these is finding a suitable student. As many of these projects sound specific and quite difficult, possibly being worded in jargon which is not current in statistics, it can be difficult to find a student. On occasions detailed investigations shows that the problem is not there! On other occasions the original industrial contact vanishes in between the signing-up of the student and the student starting. Similarly, on occasions, the industrial goes 'busy' and then later comes up with some vital information which was required at a much earlier stage.

OTHER FORMS OF COLLABORATION

Some other forms of collaboration with non-academic contacts that have been undertaken are those in which the collaborating body has financed one or two members of the lecturing staff to undertake specific research tasks. Two in particular are

(a) Cluster analysis of the elderly population of England.

(b) The household Formation Project.

Three members of staff were involved in (a) and in fact the original project which was with the Department of the Environment (DOE) was extended and further financed by the Department of Health and Social Security (DHSS) and the DOE. One outcome of this work was a number of published papers not only in statistical journals but in those aimed directly at workers in the field.

In project (b) two members of staff and two research associates were contracted by the DHSS to consider how social and economic information might be used to forecast numbers of households.

Further contract work currently underway involves

(i) I.C.I. Mond Division - development of an interactive user-friendly computer program for the analysis of regression data.

(ii) The Overseas Development Administration - development of a user-friendly survey analysis package for the micro-computer.

(iii) The D.H.S.S. - cluster analysis of the English Personal Social Services Authorities.

The above projects have been organised by the Applied Statistics Research Unit at Kent which has been set up specifically to undertake contract work from outside bodies. The Unit, which is run by a Director and an Assistant Director, employs a number of staff to undertake the routine statistical work. The pharmaceutical industry in particular has made wide use of the Unit's facilities. Members of the lecturing staff are also involved in the Unit's work either in a consultancy role or directly contracted to undertake statistical work.

A two-year project on intercropping led by Professor S.C. Pearce and financed by the Overseas Development Administration was also recently completed at Kent.

Another form of collaboration is through the Collaborative Training Award scheme financed by the S.E.R.C. At the present time one student is working on the Papadakis method in collaboration with Shell Research at Sittingbourne, Kent, and in October a student begins work on normality testing for Smith, Kline and French Ltd. at Welwyn Garden City, Hertfordshire.

SOME FINAL REMARKS

Some concluding remarks are as follows:

The solution of practical problems so often comes up with the need for theoretical research. Examples of this can be given very easily from the examples given above and references given below.

Some industries are severely put off by their contact with universities. Attempts by universities to build bridges on their own terms will do more damage to this cause than anything else. Some people are interested in bridges with the real world provided, for example, the real world is interested in their particular stochastic process. Alternatively the original problem posed by the industry is diverted into something of more interest academically. One has to realise that for the industrial statistician he has a problem and he wants it solved - anything else is, for him, a pure by-product.

References

[1] Bassett, E.E., Clewer, A., Gilbert, P. and Morgan, B.J.T. (1980) *Forecasting Numbers of Households: The Value of Social and Economic Information*, Report - U.K.C.

[2] Bravo, P.C. (1980) Sampling inspection by variables with known variance. *Bulletin in Applied Statistics*, Vol.7, No.2, pp.203-222.

[3] Bremner, J.M. (1978) Algorithm AS 123. Mixtures of beta distributions. *Applied Statistics*, 27, pp.104-109.

[4] Jolliffe, I.T. (1980) Multivariate analysis and weather forecasting. *Probabilistic and Statistical Methods in Weather Forecasting*, pp.103-108, World Meteorological Organisation.

[5] —— (1979) How accurate are long-range weather forecasts? *New Scientist*, No.84, pp.192-194.

[6] Jolliffe, I.T., Jones, B. and Morgan, B.J.T. (1982) Utilizing clusters: a case study involving the elderly. *Journal of the Royal Statistical Society*, Series A, Vol.145, pp.224-236.

[7] Jones, B. (with J.A. Eccleston) (1980) Exchange and interchange procedures to search for optimal designs. *Journal of the Royal Statistical Society*, Series B, Vol.42, pp.238-243.

[8] —— (1980) Exchange and interchange procedures to search for optimal row-and-column designs. *Journal of the royal Statistical Society*, Series B, Vol.42, pp.372-376.

[9] Jones, B. (1980) Combining two component designs to form a row-and-column design. *Applied Statistics*, Vol.29, pp.334-337.

[10] Jones, B. (1980) The Computer as a Statistical Consultant. *Bulletin in Applied Statistics*, Vol.7, No.2, pp.168-195.

[11] Köllerström, J. (with T.C. Tozer) (1980) Penalties of hard decision in signal detection. *Electronics Letters*, No.16, pp.199-200.

[12] Köllerström, J. and Wetherill, G.B. (1981) Sampling by variables with a normal-gamma prior. In: H.-J. Lenz, G.B. Wetherill and P.-Th. Wilrich (eds) *Frontiers in Statistical Quality Control*, Physica-Verlag.

[13] Morgan, B.J.T. and Robertson, C.C. (1980) Short-term memory models for choice behaviour. *Journal of Mathematical Psychology*, Vol.21, pp.30-52.

[14] Morgan, B.J.T. and Watts, S.A. (1980) On modelling microbial infections. *Biometrics*, Vol.36, pp.317-321.

[15] Morgan, B.J.T. and North, P.M. (1980) On using cluster-analysis and multi-dimensional scaling for describing animal movements and bird territories. *Biometrical Journal*, Vol.22, pp.525-533.

[16] North, P.M. and Morgan, B.J.T. (1979) Modelling Heron survival using weather data. *Biometrics*, Vol.35, pp.667-681.

[17] North, P.M. (1977) A novel clustering method for estimating numbers of bird territories. *Applied Statistics*, Vol.26, pp.149-155.

[18] North, P.M. (1981) Models for a simple capture-recapture procedure to estimate population size. *Biometrics*, Vol.37, pp.661-672.

[19] North, P.M. (with R.M. Cormack) (1981) On Seber's method for estimating age-specific bird survival rates from ringing recoveries. *Biometrics*, Vol.37, pp.103-112.

[20] Paul, S.R. (1979) A clumped beta-binomial model for the analysis of clustered attribute data. *Biometrics*, Vol.35, No.4, pp.821-824.

[21] Paul, S.R. (1979) Models and estimation procedures for the calibration of examiners. *British Journal of Mathematical and Statistical Psychology*, Vol.32, pp.242-251.

[22] Pearce, S.C. (with B. Gulliver) (1979) Graphical assessment of intercropping methods. *Journal of Agricultural Science*, Vol.93, pp.51-58.

[23] Pearce, S.C. (1980) Randomized blocks and some alternatives: a study in tropical conditions. *Tropical Agriculture* (Trinidad), Vol.57, pp.1-10.

[24] Pearce, S.C. *The agricultural field experiment: a statistical examination of theory and practice*. John Wiley, Chichester.

[25] Robertson, C. (1982) Modelling recall order in free recall experiments. *British Journal of Mathematical and Statistical Psychology* (to appear).

[26] Wetherill, G.B. and Köllerström, J. (1979) SPRT's for the normal correlation coefficient. *Journal of the American Statistical Association*, Vol.74, No.368, pp.815-821.

[27] Wetherill, G.B. (with P.C. Bravo) (1980) The matching of sampling plans and the design of double sampling plans. *Journal of the Royal Statistical Society*, Series A, Vol.143, Part 1, pp.49-67.

[28] Wetherill, G.B. (with H.-J. Lenz and P.Th-Wilrich) (1981) Editors of *Frontiers in Statistical Quality Control*, Physica-Verlag.

[29] Wetherill, G.B. et al. *Advanced Regression Analysis* (in preparation).

CLOSING THE GAP BETWEEN STATISTICIANS AND STATISTICS USERS

Guido E Del Pino and Carlos Prado-Campos, Catholic University of Chile*

SUMMARY

In this paper we discuss :

a) The present state of the teaching of statistics in Chile, with particular emphasis on the Catholic University of Chile at Santiago (U.C.).

b) Our experience with attending the demands of several users.

c) Some possible solutions to close the gap between statisticians and statistics users.

We also offer some comments on the importance of closing the gap mentioned in c).

INTRODUCTION

Closing the gap between statisticians and experts in other fields is very important on at least two counts. On the one hand, major statistical developments have been made by the users themselves. This is the case, for instance, with psychologists and multivariate analysis, economists and Bayesian analysis, astronomers and variance components analysis, to mention only a few. On the other hand, if statisticians are not acquainted with some field of applications, statistics becomes something too theorical to be of any use. Since statistics is a relatively new science, there tends to be a natural gap between statistics and the needs of its users.

It is then not surprising that important statistical associations like the International Statistical Institute (ISI), the American Statistical Association (ASA) and the Inter-American Statistical Institute (IASI) have been seriously concerned with the study of this problem through several meetings, such as Conference on Transfer of Methodology between Academic and Government Statisticians, Reston, Virginia, USA (ASA), the Latin-American Seminar on the Teaching of Statistics (SELENES) Santiago, Chile, 1977 (IASI) and the Round Table Discussion on the University Teaching of Statistics in Developing Countries, The Hague, Holland, 1968.

THE TEACHING OF STATISTICS IN CHILE

A prominent role in the teaching of statistics in Chile has been played by the Inter-American Statistical Training Center (CIENES) and by its predecessor, the Inter-American Center for Financial Statistics (CIEF) that was established in 1952, both organisations being dependent on the Organization of American States (OAS).

It is a fact that there is no university in Chile with a statistics program whose faculty does not have a graduate from CIENES. On top of this, it is also true that most Chilean statisticians with a Ph.D. degree have been previously on the CIENES programme.

Undergraduate statistical programmes offered by the universities were initially three years long and they led to the professional degree of Statistical Technician. This situation has now changed and at present some of the programmes offered are at least four years long. The universities with programmes in Statistics are shown in exhibit 1.

The curricula of the undergraduate programmes offered by the Catholic University of Chile and by the Catholic University of Valparaiso are shown in Appendix 1.

THE TEACHING OF STATISTICS AT THE CATHOLIC UNIVERSITY OF CHILE

The first step was taken in 1971 with the creation of a professional programme of statistical technician, with a length of three years. The aim of this programme was to prepare a professional to support users undertaking studies that needed statistical methodology. For this reason the curriculum included courses in sampling theory, analysis of variance, computer science and quality control.

This programme was in force for five years. During this period the need was detected for producing a professional of higher category, able to assume bigger responsibilities and consequently earning a higher income. For this reason the programme was replaced by a four year programme leading to the degree of statistician, which incorporated a broader range of courses of formative value.

The results otbained to date are satisfactory, since ninety per cent of the graduates have reasonable jobs in different firms, public and private institutions and research institutes.

It was felt, however, that it was necessary for students to have a stronger background in basic sciences and a new five year programme was created leading to the degree of Licentiate in Mathematics with major in Statistics and to a new professional degree of Statistician. The new programme includes a one semester consulting course in statistics and a one semester professional practice course.

UNIVERSITY	CITY	ACADEMIC OR PROFESSIONAL DEGREE	LENGTH
Iquique	Iquique	Statistician	4
Valparaiso	Valparaiso	Licentiate in Statistics	5
Catolica de Valparaiso	Valparaiso	Licentiate in Statistics	4
		Magister in Statistics	2
Catolica de Chile	Santiago	Licentiate in Mathematics with Major in Statistics	5
		Doctor in Mathematics with Major in Statistics	4
Chile	Santiago	Mathematical Engineer with Major in Statistics	6
Concepcion	Concepcion	Magister in Statistics	2*
Austral de Chile	Valdivia	Statistician	4

* Graduate programme

EXHIBIT 1

The curricula of the programmes mentioned are presented in Appendix 1. Some of the professional practices entered in different institutions are presented in Appendix 2.

EXPERIENCE WITH USERS

There are two basic classes of problems that typically arise in contacts with users of statistics.

a) Lack of a suitable common language to establish good communication.

b) Application of statistical methodology with disregard of the basic assumptions.

The first problem is, doubtless, mainly responsible for the break between statisticians and users. Although the second problem is solvable, if it becomes systematic it brings discredit to the statistical profession. One should point out that these problems are foreseeable. As T.A. Bancroft says "the present advance and complexity of sciences makes it difficult for the specialist to keep up to date in his own field. For this reason it is virtually impossible for him to be aware of the advances in statistical methodology".

SOME SOLUTIONS TO CLOSE THE GAP BETWEEN STATISTICIAN AND THE STATISTICS USER

To close the gap between statisticians and statistics users we make the following suggestions:

a) To offer graduate programmes directed at professionals from different disciplines.

b) To offer short courses on statistical topics of general interest, or on recent advances in statistical methodology.

c) To give a statistical consulting service to the rest of the university and outside users. This must be connected to a statistical consulting course for students in statistics.

d) To give students in statistics in-depth experience of the language and concepts of a field of application. After learning this first language, it is felt that the students would, easily learn the language of a second field of application.

e) To strengthen the role of National Statistical Associations.

f) To keep in close contact with graduates of statistical programmes. Their experience should be of great use in updating the curriculum and at the same time they would benefit by learning recent advances in statistics.

g) The statistics group in the university should offer service courses to all disciplines.

h) To illustrate methodological courses with real life examples brought from different fields of application.

i) To keep close contacts with information collecting organizations, especially with National Statistical Institutes. One way to do this is through the periodic realization of the professional practices of statistics students.

APPENDIX I

CURRICULUM OF THE LICENTIATE IN MATHEMATICS WITH MAJOR IN STATISTICS AT THE CATHOLIC UNIVERSITY OF CHILE

SEMESTER I

	Number of Credits
Algebra and Introduction to Calculus	12
Geometry I	12
Introduction to Computer Sciences I	10
Introduction to Probability and Statistics	10

SEMESTER II

Calculus I	12
Geometry II	12
Computer Sciences II	15
General Studies	5

SEMESTER III

Calculus II	8
Linear Algebra I	10
Introduction to Foundations of Mathematics	10
Statistical Methods	14
General Studies	5

SEMESTER IV

Calculus III	12
Differential Equations	8
Probability I	10
Linear Algebra	10
Introduction to Numerical Algorithms	5

SEMESTER V

Deal Analysis	16
Probability II	16
Regression Analysis	16

SEMESTER VI

Statistical Inference	16
Analysis of Variance	16
Sampling Theory	8
Numerical and Optimization Methods	10

SEMESTER VII

Stochastic Processes	14
Introduction to Multivariate Analysis	12
General Studies	5
Optional Courses	20

SEMESTER VIII	Number of Credits
Consulting Laboratory	15
General Studies	5
Optional Courses	30

SEMESTER IX	
Professional Practice	20
General Studies	5
Optional Courses	20
TOTAL	424

LIST OF OPTIONAL COURSES IN STATISTICS

1. General Methods

Time Series Analysis	12
Discrete Variable Analysis	12
Non-parametric Methods	12
Design of Experiments I	12
Design of Experiments II	12
Data Analysis	10
Topics in Statistical Methods	10

2. Statistical Theory

Foundations of Statistical Inference	10
Bayesian Inference and Decision Theory	10
Statistics Seminar	10

3. Specialized Statistical Methods

Statistical Methods in Econometrics	12
Statistical Quality Control	10
Introduction to Biostatistics	10
Topics in Biostatistics	10

4. Probability theory and Stochastic Processes 12

Stochastic Processes II
Topics in Probability Theory

5. Applied Probability Models

Reliability theory	10
Queuing Theory	10
Inventory analysis	10
Stochastic Processes Applied to Biology and Health	10

6. Statistical Computing

Simulation	10
Computational Methods in Statistics	10
Workshop on Statistical Software	10

CURRICULUM OF THE DEGREE OF STATISTICIAN AT THE CATHOLIC UNIVERSITY OF VALPARAISO

SEMESTER I	Number of Credits
Algebra	6
Calculus I	6
Elements of Statistics I	6
Physical Education	1
Religious Culture I	2
SEMESTER II	
Linear Algebra	4
Calculus II A	4
Elements of Statistics II	8
Physical Education II	1
Religious Culture II	2
SEMESTER III	
Complements of Linear Algebra	4
Calculus II B	4
Probability Theory	6
Survey Design	3
Physical Education III	1
Religious Culture III	2
SEMESTER IV	
Elements of Numerical Analysis	4
Complements of Calculus	6
Statistical Inference	6
General Studies	2
SEMESTER V	
Operations Research	6
Sampling Theory	6
Optional Course	4
General Studies	2
SEMESTER VI	
Experimental Design	6
Optional Course	4
Optional Course	4
Optional Course	4
SEMESTER VII	
Statistical Workshop I	6
Regression Analysis	6
Optional Course	4
SEMESTER VIII	
Statistical Workshop II	6
Statistics Seminar	6
General Studies	3
TOTAL	145

A P P E N D I X II

REPORTS ON THE PROFESSIONAL PRACTICES MADE BY STUDENTS OF STATISTICS OF THE CATHOLIC UNIVERSITY OF CHILE

In the following list are registered the name of the professional practices entered by the students and the place where the work was undertaken.

	NAME	PLACE
1.	Fit of a nonlinear regression model to an agronomic problem.	School of Agriculture, U.C.
2.	An application of statistics to forestry research.	National Forestry Research
3.	Statistical analysis of private undergraduate educations programmes.	National Statistical Institute
4.	Statistical analysis of movie preferences in Santiago.	Compañia distribuidora de películas (CONATE)
5.	Statistical applications data processing.	National Electricity Company
6.	Marketing research on powder milk.	CRAVAL
7.	Study of the academic efficiency of Psychology students of the Universidad Catolica.	School of Psychology of the U.C.
8.	Statistical use of a data bank.	Las Viscachas Company.
9.	An application of statistics to agronomy research.	National Institute of Agronomy Research
10.	An application of statistics to quality control.	CTI
11.	Estimation of the mean number of failures in electronic instruments.	Subway of Santiago
12.	Market research in fabrics.	Yarur S.A.
13.	Efficiency of looms in the Yarur company.	Yarur S.A.
14.	Distribution of compensations and other guarantees in bakery workers.	Departamento de Indemnizacion de obreros molineros y panificadores
15.	Evaluation of a quality control system.	Phillips Chilena
16.	Estimation of the long distance call income.	Chilean Telephone Company
17.	Level of fertility research in Chile.	Latin American Center of Demography
18.	Inventory control.	Chilean Telephone Company
19.	Application of PERT analysis to geological problems.	Geology Institute

20.	Applications of different statistical methods in agronomy research.	National Institute of Agronomy Research
21.	Research on the trend of mortality by sex and age.	Latin American Center of Demography
22.	Sampling and checking an Inventory in the Andina Mining Company.	Andina Mining Company
23.	Time series analysis of the demand for mushrooms.	Pro-Chile
24.	Forecasts of illiteracy (1975-2000).	National Statistical Institute
25.	Error analysis in determining the laws of copper and molibdenum.	Research Center for Mining and Metallurgy
26.	Forecast of automobile demand between 1980 and 1989.	Banco Hipotecario de fomento
27.	Quality control of raw materials in a pharmaceutical industry	Schering Laboratory
28.	Portfolio analysis	U.C. Computer Center
29.	Analysis of lattice designs.	National Seeds Company
30.	Market research in housing.	Colocadora National de Valores
31.	Forecasting textile imports.	Textile Institute of Chile
32.	Analysis of automobile accidents	BHC Insurance Co.
33.	analysis of malnutrition in Chile.	CONPIN
34.	Study of the quality control of services of 3M-chile.	3M-Chile
35.	Forecasting models for food imports.	SOFOFA
36.	Study on the consumption of fertilizers and pesticides in Chile.	computing Promotional Co.

CO-OPERATION BETWEEN ACADEMIC AND PRACTISING STATISTICIANS

A Prat Bartés, AVIA, Barcelona

The main purpose of this contribution is to illustrate some of the benefits that co-operation between our Department of Statistics at the Polytechnic University of Barcelona and the public administration and/or private industrial sector in Barcelona, Spain, has produced.

As George Box (1980) pointed out, some discoveries of the stature of least squares, ratio estimators, design of experiments and others, have been the result of the interaction with reality. Our Department is firmly convinced that co-operation with the industrial environment in real problem solving is very important in order to stimulate theoretical research.

Our involvement with real problem solving has been possible in part due to the engineering background of most of the personnel in our Department. Knowledge of the physics of many industrial systems and familiarity with computers have proved very useful tools when used in connection with statistical theory and methods.

Some examples of how reality has stimulated theoretical research in our Department are the following:

1. During our co-operation with Electricity Utilities of the Barcelona area in a research project about load curves, a problem of non-linearity in the relationship between temperature and daily electricity consumption arose. This has stimulated a Ph.D. research project on the topic of non-linearity in transfer function models.
2. When co-operating with a banking organisation in Barcelona with some 1,000 branches we found that it would be almost impossible to forecast the deposits of each of the branches using the Box-Jenkins approach. The need for a reliable automatic forecasting system arose.

One of our students, Moisés Valls, has almost completed his Ph.D. work which gives a solution to the problem of automatic modelling by defining a statistical distance between theoretical autocorrelation functions and inverse autocorrelation function, and the corresponding sample functions. The method seems to work very well with real data.

There are many other benefits arising from this co-operation, especially in getting good material for teaching. With the permission of the Electricity Utilities we use a very important data-base obtained during the research project as real data for teaching purposes. At the beginning of the semester, each student gets a portion of the data-base and all students are trained in using a statistical package (SPSS in our case). Most of the general statistical course is based on this real data set. A great number of statistical methods can be used here: checking the quality of the data, exploratory data analysis, linear models, mechanistic and empirical model building, time series, influential observations, etc. Using real data increases the motivation of our students to learn the theory necessary to understand the methods required to analyse those data.

Furthermore, as Ron Snee pointed out in his talk here at ICOTS, through this co-operation with local industry we have been able to use companies' big computers that were not available to us in our University at this time.

Another area of co-operation with local industry has been the training of industrial statisticians in the most recent developments of our subject matter, through short courses designed specifically for this purpose. Some of these short courses have been given "in house" to the industrial practitioners, and others have been lectured in our Department.

Those one-week courses have covered subjects such as linear models, time series analysis and survey sampling.

When dealing with small and medium size industries our efforts have not been very successful yet. Our Department offers short courses in Statistics through the CPI (Center for Continuous Education of the Catalan Engineering Association) that are attended by some of the engineers working in these types of industries. We are developing statistical software tailored to the needs of these industries and specifically designed for minicomputers. Our feeling is that much effort should be developed in this area.

As a way of summarizing this talk we would say that the interaction between reality and theory in our Department coming from our co-operation with local industries and public administration, has been extremely fruitful from the point of view of both academic and industrial institutions. It has provided real problems to be solved through our research programme and excellent material for teaching.

The industries, on the other hand, have been introduced to the most recent methods of statistics and in some cases have experienced great reductions in production costs through a sound use of data analysis.

Finally, we feel that more effort should be deployed in the co-operation with small and medium size industries.

Reference

[1] Box, G.E.P. (1980) Comment on the paper: "Preparing Statisticians for Careers in Industry: Report of the ASA Section on Statistical Education Committee on Training of Statisticians for Industry". American Statistician, 34, 2, 76.

TOPIC 12: TEACHING SURVEY SAMPLING

ORGANISER: DR D SINGH

THE TEACHING OF SAMPLE DESIGN TO GOVERNMENT STATISTICIANS FROM DEVELOPING COUNTRIES

Joseph K Garrett, U S Bureau of the Census, Washington, D C

INTRODUCTION

The International Statistical Programs Center (ISPC) of the U.S. Bureau of the Census has conducted training programs for government statisticians from developing countries for more than 30 years. Through these programs, ISPC has contributed substantially to statistical development in many countries. The programs are conducted under the auspices of the Agency for International Development and serve urgent and changing needs of developing countries for trained personnel to collect, process, and analyze statistical data. Currently, 1-year training programs are offered in each of the following specializations:

(1) Sampling and Statistical Methods
(2) Computer Data Systems
(3) Survey Methods
(4) Agricultural Statistics
(5) Demographic Statistics
(6) Economic Statistics

These six programs are directed toward training statisticians in the specific skills needed to carry out many different phases of ongoing statistical activities involved primarily in conducting socioeconomic household surveys, agricultural surveys, and retail establishment surveys. Although participants specialize in one program, they are introduced to clearly related and auxiliary fields and skills to give them an understanding of how their specialty is linked with the total statistical system.

This paper discusses the education and training given to participants specializing in Sampling and Statistical Methods at the ISPC. This education and training in survey sampling is conducted through linkage of sampling concepts with the various other survey-related skills and fields. This means, among other things, that it is not sufficient to merely offer a course on sampling theory when the objective is to train participants to be functioning sampling practitioners and technicians. While this paper primarily discusses the training in sampling conducted at the ISPC for statisticians from developing countries, it is nevertheless felt that most of the remarks are relevant to the teaching of survey sampling in general.

THE CONCEPT OF LINKAGE

It would seem self-evident to assert that when teaching any subject, or when training in any subject-matter area, that subject needs to be linked to related disciplines and auxiliary fields. While this appears obvious, the concept of linkage is often absent in traditional courses on survey sampling. More often than not, the sampling course becomes a course in algebra. Algebraic results showing an estimator to be unbiased or of smaller variance than that of a competing estimator seem to be the dominating feature of many class lectures and homework problems (Murthy, 1980). While such results are indeed important and should be included in sampling courses, too much emphasis on algebra at the expense of linkage tends to create "conditional samplers". That is, students of such courses may become adept at solving problems such as: "Given the simple random sample $x_1, x_2, \ldots, x_n$, show that $E(\bar{x})=\bar{X}$ and that $V(\bar{x})=(1-f)\frac{S^2}{n}$ ". While developing a student's ability in this type of algebraic endeavor is a necessary exercise in order to build his or her understanding of the mathematical-statistical framework upon which population sampling is grounded, its overemphasis would tend to ignore the fundamental problem of how a simple random sample of n observations is actually obtained. To the extent that this conditional approach dominates the more complicated (and most often used) topics of stratified, clustered, and multi-stage sampling, the student is not suitably prepared to deal with the applied issues of data collection that are involved in survey sampling. To such students, the concept of a primary sampling unit may be nothing more than an algebraic entity.

In teaching survey sampling, one must not only develop the mathematics of survey sampling, but also provide for the integration of sampling topics with the various other important survey-related topics. Development of a survey sampling statistician requires linkage to be developed in the context of the overall objectives of the training program.

THE DEVELOPMENT OF LINKAGE

In order to develop the concept of linkage in the training of survey sampling to statisticians from developing countries, two broad tasks need to be accomplished. First, fields and skills clearly related and auxiliary to survey sampling need to be identified. These include, among others, design of tables and questionnaires, editing and coding principles, methods of imputation, data processing, survey budgeting and scheduling, survey management, control and evaluation of nonsampling error, geography and mapping, statistical quality control, enumerator recruitment and training, and presentation and publication of data. Second, an assessment of the general survey conditions pertinent to survey sampling in developing countries needs to be done. These two tasks are not entirely mutually exclusive. Rather, in assessing the general survey conditions, some needed skills clearly related to survey sampling in developing countries are identified. Moreover, such assessment provides the basis for emphasis that certain topics receive during the training program. While it is not the purpose of this paper to fully characterize the general survey conditions of the developing countries of the world, it is possible to list a few that are inherent in many developing countries:

- Available maps are often inadequate or outdated.
- Transportation in many parts of the country is difficult; good roads are few and far between.
- Complete frames which can be used for sampling arenot available; frame construction is an essential.
- Independent sources of data are skimpy and often unreliable.
- Nonsampling error is often the dominant source of error in survey estimates.
- Survey budgets are quite limited and managed, at times, in a haphazard manner.
- Overall quality control is poor and often not regarded as important.
- The "sampler" does not havedesigning samples as his/her only job. He/she must not only have specific sampling knowledge, but also general survey knowledge.
- Subject-matter concepts are often poorly or inconsistently defined.
- Training of enumerators is generally not of sufficient quality.
- Communication among various survey personnel can be improved.
- Housing stock is generally poor and living conditions are quite crowded.

Using this fairly broad assessment of general survey conditions in developing countries in conjunction with identification of clearly related and auxiliary fields and skills, the concept of linkage can be developed. Not only is there the pedagogical need to link the concepts of sample design with other important survey-related activities in the actual teaching process, but participants should also be introducd to the various other survey-related activities so as to provide a meaningful setting for the linkage. For example, in order to link the sampling concept of "determination of sample size" with the survey concept of "output tables", participants need to have a conceptual grasp on what "output tables" are.

At ISPC, the development of linkage in the training of survey sampling to government statisticians from developing countries is accomplished on two levels. In the actual teaching of sampling concepts, frequent reference is made to the other important survey-related activities. In addition, to cement the concept of linkage, education and training in these other important survey-related activities is also conducted concurrently. This is all done through classroom lecture and discussion, "learning by doing" laboratories, and case study workshops. In addition, a large-scale demonstration survey is planned, designed, managed, and ultimately conducted in the field by the participants.

While emphasis in this paper is on the application of the concept of linkage of the sampling concepts to the various other survey-related concepts, a description of a typical training plan for ISPC sampling participants is given in Appendix A.

APPLICATION OF LINKAGE IN THE TEACHING OF SAMPLE DESIGN

To illustrate linkage in the teaching of sample design to government statisticians from developing countries, several examples will be given. These examples are primarily concerned with socioeconomic household surveys and agricultural surveys conducted in developing countries and consequently emphasize general survey conditions present in many of these countries. However, several of the remarks are relevant for surveys conducted in developed countries as well. On a more general level, the application of linkage of sampling concepts to related survey topics is a principle applicable in varying degrees to any course on survey sampling. The examples that follow focus on important and basic topics such as survey error, determination of sample size, stratification, clustering, multi-stage sampling, area sampling, point estimation, and variance estimation. While these do not exhaust survey sampling topics, they suffice to illustrate the approach by which sampling topics can be integrated with and linked to other survey-related activities in sample design courses.

Survey Error: Sampling error is, quite obviously, an extremely important concept in survey sampling. Controlling sampling error to a tolerable level is a required design feature. In many developing countries, however, nonsampling error is often the dominant source of error in survey estimates. Thus, in teaching sampling to developing country statisticians, it is better to introduce from the beginning a broad framework for sample surveys which includes both sampling and nonsampling errors (Hansen, Hurwitz and Bershad, 1961; Murthy, 1980). This introduces the notion of "total error" of a survey estimate. For discussing nonsampling error, this author has used a framework which classifies sources of nonsampling error by the respective survey stage that the source initially appears (Garrett, 1979). Thus, sources of nonsampling error are found in the preparatory stages of surveys, the data collection stage, the data processing stage, and the publication stage. This framework serves two purposes. It provides a method for introducing a concept which by its very nature is diverse and pervasive. Moreover, it serves as an approach to discuss controlling nonsampling error throughout the survey cycle. Preparatory stage errors can occur due to problems of specifying survey goals and objectives, improper planning and scheduling of survey operations, use or development of inadequate maps, use or development of faulty frames, and inadequate recruitment and training of enumerators. Errors in the data collection stage include errors of coverage, content, and nonresponse. Errors in the data processing stage include errors of editing, coding, imputation, key punching, computer programming, tabulation, and calculation. Publication stage errors include inconsistent tables, statistically inaccurate statements, and misleading narrative.

Insofar as the first step toward controlling nonsampling error is identification of its sources, the above outline indicates many of the more common sources of nonsampling error in surveys planned and conducted in developing countries. The concept of controlling sampling error needs to be linked to the concept of controlling nonsampling error so as to establish control of total error. Thus, the concept of nonsampling error should be developed concurrently with that of sampling error and discussions related to the control of nonsampling error should point out relevant survey procedures, methods, and activities that provide such control. These include necessary improvements in mapping, questionnaire and table design, planning of field operations, and overall survey quality control.

Determination of Sample Size: The sampling problem of determining an appropriate sample size should be linked with the survey concept of output tables. Reviewing intended survey output tables gives both the survey sponsor and the sampling technician a feeling for the magnitude of the proposed survey (Ford and Tortora, 1978). Throughout such review, attention can be directed toward the desired reliability requirements for the various major tables and cells within. Additionally, review of output tables helps in determining possible stratification criteria such as "urban" and "rural". As estimates of sample sizes for major characteristics of various table grand estimates, marginal estimates, and cellular estimates are made, it may be found that certain cells or tables need to be collapsed. Also, it may be found that the survey is overloaded with major characteristics which are too diverse. In many developing countries, survey budgets are quite limited and management of such budgets can be improved if proper attention is given to careful review of output tables in connection with determining sample size. Available data for estimating variances required for estimating sample sizes is skimpy or of poor quality in many developing countries and crude estimates often need to be resorted to. Rough estimates of variance can be calculated using information on the approximate range and shape of the distribution of the major survey characteristics (Deming, 1960) and these approximation methods need to be stressed in class.

Stratification: The topic of stratification offers many opportunities to introduce clearly related survey activities. While many texts emphasize size stratification, geographic stratification is often used for multipurpose surveys in developing countries. Topographic maps are often used in conjunction with subject-matter knowledge in developing stratification for, say, an agricultural survey. So-called "land use strata" are delineated on maps when defining strata to be used with area sampling. Satellite imagery, which is available for most countries, is often useful for stratification. For household surveys, geographic stratification makes sense if major household survey characteristics are similar within broad socioeconomic neighborhoods or regions. In the teaching of the principles of geographic stratification, laboratory exercises, case-study workshops, and demonstration surveys are used at ISPC to allow sampling participants the opportunity for "hands-on" experience in developing strata. These exercises emphasize that when delineating stratum boundaries, use of natural physical features which can be easily identified in the field is a primary consideration. In many developing countries, lack of quality data generally prohibits an extremely rigorous quantitative search for stratifying variables and, hence, geographic stratification is often a reasonable alternative. Thus, the topic of stratification can be linked with availability of quality maps, aerial photographs, satellite imagery, and the existence of adequate census data.

Clustering, Multi-Stage Sampling, Area Sampling: The sampling topics of clustering, multi-stage sampling, and area sampling are topics which, quite obviously, involve mathematics. In conjunction with stratification, these topics are also very applied notions which have definite relations to other survey activities. These relationships deserve proper recognition during the teaching of survey sampling. For socioeconomic household surveys or agricultural surveys in developing as well as developed countries, the general sample design is often a stratified, clustered design involving more than one stage of sample selection. Often, the first-stage sampling units are geographic areas. While the necessary algebra for this general design needs to be developed, it is also the situation that these topics need to be linked with a variety of other survey activities. The topic of clustering basically involves the choice of a suitable sampling unit. It should be stressed that existing administrative units should be reviewed to see if any would also be appropriate sampling units. For national socioeconomic household surveys this review may provide primary sampling units which are suitable from not only a statisical viewpoint, but also from a cost-benefit approach. For agricultural surveys, nonadministraive geographic areas delineated on topographic maps (if such are available) within land use strata may represent reasonable primary units. Thus, complementary to the mathematics, an applied, spatial notion of "primary sampling unit" needs to be developed. This can often be done in teaching by relating the concept of "primary sampling unit" with existing administrative units and available maps.

Generally, subsampling within primary units is necessary due to budgetary and staffing constraints. This is especially true in developing countries. During teaching, the algebra of subsampling needs to be balanced with the details of this process in the field. The survey topics of map preparation for selected primary units, segmentation of primary units, field listing and screening, and enumerator recruitment and training should be linked with the sampling topic of multi-stage sampling. Additionally, emphasis should be given to the control of nonsampling error during these field operations. If second (and further) stage sampling units are geographic areas, the use of physical boundaries in the mapping process defining such units needs to be emphasized.

Area sampling in which completely enumerated second-stage sampling units are geographic areas within selected first-stage area units is a commonly used design for agricultural surveys. The methodology involved during data collection with this design has impact on questionnaire wording and is a function of the characteristics being measured. For example, crop information is generally gathered for the entirety of land area within sample segments operated by farmers living either in or out of the sample segment. However, information on, say, cattle is collected only from farm operators residing within the sample segments. For multi-purpose agricultural surveys of this design, the sample design is closely related to design of the questionnaire. Failure to observe this relationship may result in estimation problems once the data are collected. Thus, relationship between sample design and questionnaire design needs to be observed when conducting sampling courses.

The topics of clustering, multi-stage sampling, and area sampling should have the requisite mathematics developed while concurrently developing the linkage of these topics to geography and mapping activities, review of available administrative areas which might serve as statistical units, details of various field operations, enumerator recruitment and training, questionnaire design, and survey budgeting.

Point Estimation, Variance Estimation: The survey activities that occur after data are collected need to be emphasized. It is not a good portrayal of survey reality to discuss estimation of various population parameters abstracted from such important survey activities as editing, coding, keypunching, adjustments for non-response, and other data "cleaning" and data preparation activities. To discuss point estimation as simply a parochial and clinical algebraic process does not suffice. The sampling technician in developing countries is involved, directly or indirectly, in all of these activities. During the teaching process, frequent reference should be made to the importance of these activities. The usefulness of the estimating formulae in sampling can be linked directly to the quality with which these data processing activities are completed.

After data preparation activities are completed, the sample data needs to be "weighted" so as to obtain population estimates. In survey practice, full communication between the sampling technician and the data processing specialist needs to be realized. Communication among survey personnel can be a problem in developed as well as developing countries. It is the sampling technician's responsibility to provide lucid specifications precisely describing how this weighting should be done - including all necessary adjustments to the basic weight derived from the selection probabilities. For large-scale national surveys, the sampling concept of "self-weighting design" can be readily linked to this possible communication problem. In the sampling course, development of weighting specifications needs proper consideration and involves much more than simply describing the algebraic form of the various estimators.

For large-scale, multi-purpose surveys involving an extraordinarily large publication stub, variance estimation is a complex undertaking. For national-level surveys involving numerous characteristics, adjustments of various kinds for nonresponse and undercoverage, estimation features such as ratio adjustments and large observation procedures, and design features such as stratification, multi-stage sampling, and unequal selection probabilities, the variance estimation approach needs to consider not only accuracy, but also timeliness, cost, simplicity, and other administrative considerations (Wolter, 1979). In developing countries constrained by computing facilities, survey budgets, and survey operations staff, these considerations become very important. In addition to customary (textbook) variance formulae, other variance estimation methods used in practice such as random group procedures, balanced half-sample techniques, and methods for systematic sampling need to be introduced. These ideas can also be related to report writing and the need for statistical standards in describing the findings of the survey.

REINFORCEMENT OF LINKAGE IN THE TRAINING OF SAMPLING

While constant reference to related survey topics, skills, and fields should be made during the teaching of the sampling course, it is also essential to allow students the opportunity to actually design a sample. At ISPC, the participants specializing in Sampling and Statistical Methods design either a socioeconomic household survey or an agricultural survey, depending on their interests. For both surveys, the target population is an entire county in the U.S. Working with participants specializing in other areas, these surveys are conceived, planned and ultimately conducted in the field. Beginning with definition of survey goals and objectives, proceeding through the survey cycle, and ending with "publication" of the survey results, the integrated nature of survey activities becomes apparent. In particular, the sampling participants experience for themselves that survey sampling is more than an algebraic endeavor. While the surveys are viewed as training or demonstration surveys, many real survey features are present. The socioeconomic household survey generally has a sample size of approximately 500 households selected through a stratified (urban/rural), clustered (existing administrative or statistical units), multi-stage (area segments as second-stage, listed households as third-stage) design. The agricultural survey generally utilizes a sample of area segments selected within land-use strata (intensive agriculture, extensive agriculture, forests, mountains, and urbanized) through a two-stage process. During the data collection phase of the survey, participants become enumerators and face first-hand some of the real problems involved in implementing the details of field operations. The experience emphasizes the need for a quality enumerator training program for proper data collection activities. Throughout the six-month period of involvement in the survey, participants not only perceive the various connections of sampling to other survey features, they witness the effects of ignoring relevant linkage.

The purpose of these demonstration surveys is two-fold: (1) To allow participants the opportunity to develop a technically sound sample design, complete with documented details on stratification, clustering, sample selection, and estimation; and, (2) to show the relevant and necessary linkage that binds sample design to overall survey design. For many, it may be the first time they have viewed a sample survey as an integrated process. This experience translates into better, more well-coordinated surveys when they return home.

CONCLUDING REMARKS

Sample design is basically an applied field. While probability theory and algebra are necessary ingredients in the application of sampling principles, they do not represent the entirety of sampling. In the teaching of sample design, whether it be to government statisticians from developing countries or any other group, the basic teaching problem is one of balance - balancing necessary mathematics with necessary application. Real application of sampling principles to subject-matter problems involves development, application, and maintenance of linkage of sampling concepts to other important survey skills, activities, and fields. Consequently, incorporating linkage more fully in the teaching process is necessary. It seems, as stated earlier, that this idea is obvious and self-evident. Why, then, do many sampling courses become unduly balanced by excessive algebra? I can think of two possible reasons: (1) Algebra is the most important part of sample design and should be the dominating feature, (2) experience in algebra is more fully developed in the instructor than is experience in actual application of sampling principles. If a sample design were developed to reliably estimate the proportion of instructors teaching sampling who possess the characteristic "experience in sample design application", the resulting estimate would probably be very small. Insofar as this may be true, the second reason given above may very well imply the first reason.

References

[1] Cochran, W.G. (1977) Sampling Techniques. John Wiley & Sons, New York.

[2] Deming, W.E. (1960) Sample Design in Business Research. John Wiley & Sons, New York.

[3] Des Raj (1968) Sampling Theory. McGraw Hill Book Co., New York.

[4] Deighton, R.E., Poland, Jr., et al. (1978) Statistical Policy Working Paper 4, Glossary of Nonsampling Error Terms: An Illustration of a Semantic Problem in Statistics. U.S. Department of Commerce, Washington, D.C.

[5] Ford, B.L. and Tortora, R.D. (1978) "A Consulting Aid to Sample Design". Biometrics, 34, 299-304.

[6] Garrett, J.K. (1979) "Assessing the Reliability of Data: Nonsampling Error". Paper presented at 1979 Symposium on the Analysis of Labor Statistics: Washington, D.C.

[7] Gonzalez, M., Ogus, J.L., Shapiro, G., and Tepping, B.J. (1975) "Standards for Discussion and Presentation of Errors in Survey and Census Data". Journal of the American Statistical Association, 70, No.351, Part II.

[8] Hansen, M.H., Hurwitz, W.N. and Bershad, M.A. (1961) "Measurement of Error in Census and Survey". Bulletin of International Statistical Institute, 38, (2), 359-374.

[9] Hansen, M.H.,Hurwitz, W.N. and Madow, W.G. (1953) Sample Survey Methods and Theory, Volumes I and II. John Wiley & Sons, New York.

[10] Kish, L. (1965) Survey Sampling. John Wiley & Sons, New York.

[11] Murthy, M.N. (1980) "Some Thoughts on Teaching of Sampling". Paper presented at 1980 Symposium on Teaching Statistics, Goteburg, Sweden.

[12] Wolter, K. (1979) Introduction to Variance Estimation (draft copy of text to be published).

[13] Zarkovich, S.S. (1966) Quality of Statistical Data, Food and Agricultural Organization, Rome.

APPENDIX

COURSES AND WORKSHOPS FOR PARTICIPANTS SPECIALIZING IN SAMPLING AND STATISTICAL METHODS

		Total Class Sessions
First Training Period		
104-1	Design of Sample Surveys	15
105-1	Survey Sampling Laboratory	5
600	Introduction to Design of Surveys and Censuses	15
Second Training Period		
104-2	Design of Sample Surveys - continued	30
105-2	Survey Sampling Laboratory - continued	10
601	Introduction to Data Processing for Subject-Matter Specialists	15
602	Design of Tables and Questionnaires	30
604	Budgeting, Scheduling and Controlling of Survey-Census Operations	15
	Optional:	
203-1	Agricultural Survey Operations and Methods	10
405	Applications of Economic Survey-Census Techniques	10
Third Training Period		
103	Applied Regression and Correlation Analysis	20
106	Variance Estimation in Sample Surveys	20
107	Laboratory in Variance Estimation Techniques	10
111	Introduction to the Analysis of Data	20
120-1	Field Demonstration Laboratory in Sampling and Statistical Methods	10
150-1	Seminar: Selected Topics in Sampling and Statistical Methods	10
605	Editing, Coding and Imputation Principles	30
609	Workshop: Training for Statistical Activities	1 week
695	Special Programs in Statistical Technology	10
	Optional:	
203-2	Agricultural Survey Operations and Methods - continued	10
204	Agricultural Sample Survey Design	20
205	Frame Construction for Agricultural Samples	20
430	Economic Enterprise Survey: Field Demonstration Laboratory	10
603-1	Geography and Mapping for Surveys and Censuses	20
Fourth Training Period		
108	Control and Evaluation of Nonsampling Error	20
109	Quality Control in Censuses and Surveys	10
110	Applications of Current Survey Design and Evaluation Techniques	20
120-2	Field Demonstration Laboratory in Sampling and Statistical Methods - continued	10
150-2	Seminar: Selected Topics in Sampling and Statistical Methods - continued	10
606	Management of Statistical Activities	20
608	Presentation, Publication and Dissemination of Data	10
	Optional:	
206	Objective Measurement of Area and Yield	30
308	Alternative Strategies for the Collection and Analysis of Demographic Data	10
603-2	Geography and Mapping for Surveys and Censuses - continued	20
Fifth Training Period		
190	Case Study in Sampling: Workshop and Field Demonstration	Full Time

TEACHING OF SURVEY SAMPLING

D Singh, UNDP, Amman

INTRODUCTION

Survey sampling is one of the most commonly applied disciplines of statistical science. The subject is of such a common use that even a layman without any formal training in the subject may sometime depend upon its application for solving his problems and arriving at the right decision. The decision of a consumer to purchase a bag of grain by examining only a handful of grain is one of the many examples of the application of the sampling technique by a layman. Because of the apparent simplicity of the technique, the subject of survey sampling did not receive the respectability by the mathematicians that it deserved. It was during and after the Second World War when the demand for various kinds of statistics in diverse fields of human activities for taking quick policy decisions increased, the attention of theoretical statisticians engaged in war efforts or later on in socio-economic developmental programmes was drawn to the problem of collection of data by following sampling techniques.

Immediately after the Second World War, the demand for socio-economic data became so urgent that the United Nations Sub-Committee on Statistical Sampling was set up in 1974 to mainly review the scope of sampling technique in its application to censuses and surveys. As a direct consequence of the report of this Sub-Commission, the United Nations and its specialised agencies took necessary steps for promoting the use of sampling methodology, and considerable success, particularly in developing countries has been achieved over the last three decades. In several industrialised countries, particularly in UK and USA, which already had some experience in applying sampling technique even prior to the Second World War, systematic research both in theoretical aspect as well as in its application was undertaken during the war period, and it was introduced as a part of statistics courses in a number of universities. In the Indian sub-continent, the research on the application of sampling methods for estimation of crop productions was initiated by the Indian Statistical Institute and Indian Council of Agricultural Research more than four decades ago, and later on, the method so developed was adopted by the Indian Government as a routine procedure for crop estimation.

Many developing countries in Asia, Africa and Latin America which became free from colonial rule after the Second World War started socio-economic development to fulfil the aspirations of their people. One of the major problems that these countries faced was the lack of basic information needed for the purpose. With the aid of some international agencies or friendly countries, many of them planned censuses and surveys depending upon the specific needs, and the monetary resources and manpower that they could mobilise. It was at this stage that these countries felt acute shortage of trained manpower at different levels to continue the census and survey activities on the regular basis so urgently needed. To fill up the gap of manpower at least at the intermediate level, several regional training institutes, each meeting the requirements of a group of countries of the region were established. Although these institutes impart courses in general statistics as applied to official work of national governments, an elementary course on sampling methods is an important component of the whole programme. The quality of these courses, however, depends on the quality of the teachers' and trainees' background. Many of these institutes have already functioned for more than a decade. It may be very appropriate at this stage that an independent international agency like ISI may evaluate as to how far these institutes have fulfilled the objectives for which they were established. Several national governments, besides encouraging the universities to impart courses on sampling methods as part of a degree course in statistics, established separate specialised training and research institutes. For example, in India, besides the Indian Statistical Institute, the Indian Agricultural Statistics Research Institute established immediately after the Second World War imparts advanced courses in sampling as applied to agriculture. A new Institute dealing with medical statistics recently established in New Delhi under the Indian Council for Medical Research, imparts courses on sampling as applied to medical problems. In USA, the Institute of Social Research and Statistics, University of Michigan organises special courses on sampling as applied to social sciences. The US Bureau of the Census, Washington D.C., known for its research in application of sampling techniques to census problems, organises special training programmes in sampling for the benefit of developing countries. Many such organisations came up all over the world.

With increased use of survey sampling technique, a number of books on sampling - Yates (1947, 1960), Cochran (1953, 1960), Sukhatme (1953, 1970), Hansen (1953), Murty (1967), Kish (1967), Des Raj (1968) etc. - were brought out as guides to practising statisticians and as reference literature to professional teachers. Fortunately, most of these books were written by practising statisticians and therefore, these reflect individual author's experiences. UN Statistical Office, New York, started separately monitoring information on current sample surveys conducted by the member countries, and making them available for wider use. Many national governments also bring out annual publications on the current sample surveys. All these literatures are of considerable use particularly to practising sampling statisticians.

SCOPE OF THE COURSES

All the institutions engaged in teaching of survey sampling at present may be broadly classified into three categories (i) Universities (ii) Training and Research Institutes and (iii) National Statistical Offices.

In the universities, the subject of sampling is taught as a part of the course on mathematical statistics and emphasis is mostly laid on mathematics of sampling, and derivation of various formulas relating to estimates of population parameters and their variances. Even if the subject is taught as a part of applied statistics courses, the treatment is mostly mathematically oriented because of lack of application of the teachers in the practical application of the subject. The students may successfully complete the course without developing any confidence in its application to solve the practical problems of designing a sample survey. This is one of the reasons that in many countries a large number of statistical posts in government and public undertakings remain unfilled, in spite of the fact that many university graduates in statistics are unemployed.

The training and research institutes are supposed to give a balanced course, which may consist of underlying theory of sampling and its application. This balance can be obviously maintained only if the teachers themselves are engaged in research in the development of sampling techniques and their applications to solve the practical national problems. If the need of theoretical research arises, the teachers should have the capacity and interest in undertaking such research programmes. These are the institutions where teaching, research and extension programmes are supposed to be integrated.

The training courses organised at the national statistical offices are of purely applied type. In such courses, sampling methods are mechanically taught, describing the procedure of their applications. Even if some aspects of theory are discussed, use of algebra is avoided. Participants in such training programmes come mostly from government and public undertakings which adopt sampling technique as a tool for collection of data. Sometimes, these trainees learn the full details of all stages, from formulating the questionnaires and field instructions to preparing the survey report of a specific sampling project chosen from the area of the individual's interest.

CONTENTS OF THE COURSE

It is noticed that whatever may be the level of the students, the contents of the sampling course are almost the same. Most of the teachers try to cover the topics described in a standard textbook on sampling prescribed for the course. If the course is being delivered to the postgraduate students of mathematical statistics, it is assumed that they have already covered the relevant topics on statistical methods and probability, and know adequate algebra and calculus for following the course on sampling. Descriptions on the details of the individual topics depend upon the interest of the teacher and time allotted for it. In fact the subject is taught like any other topic of a mathematical statistics course, forgetting that the sampling is basically an applied subject and relevance of the course lies in inculcating interest among the students in learning it for solving the practical problems. The result is that unless such students are given further training on application aspects, they completely lose interest in the subject and in a short period forget it. If the objective of such a course is to produce teachers and research workers, the quality of teaching and relevance of research will again depend upon the way they look to the subject of sampling.

Because of the wide use of the technique, the courses on sampling are also these days imparted to students of other disciplines, such as economics, commerce, agriculture, medicine, engineering, public administration etc. Here the objective of the course is not to train the professional statisticians but to make the students familiar with sampling methods as a scientific tool for collection of data, which may be needed in their future careers. The number of students of this category is unusually large. The topics listed in the course are mainly the same as prescribed for the students of statistics, and teachers also employed for teaching this course have mostly training in mathematical statistics without any orientation in the application of the technique to specific fields. Consequently, their teaching style and way of communication are generally mathematically oriented. They introduce the topics through algebraic notations and develop the underlying concepts of sampling through hypothetical populations. Since these categories of students generally have a poor background in mathematics, they start resisting the subject and lose interest completely in learning it further. This is the greatest harm done to the subject. These are the persons who in future become decision makers and because of lack of appreciation of what the sampling technique can do in helping them to arrive at right and quick decisions, many useful sampling projects are abandoned on the pretext of lack of financial resources. The contents of the course for such categories of students should, therefore, lay emphasis on basic philosophy of the subject, and introduction of basic concepts such as sampling and non-sampling errors, probability sampling, confidence intervals, unbiased estimate, accuracy etc. through live examples and numerical exercises.

To understand some of the concepts of sampling, training in basic statistical methods is necessary. For example, the concept of confidence interval cannot be clearly understood without proper knowledge of the properties of the normal distribution, and sampling distributions of relevant statistics. While discussing the sampling distribution of mean, the underlying idea of central limit theorem should be clearly understood, particularly for the reason that most of the survey populations are finite while the normal distribution concept is applicable to infinite and continuous population. Similarly, concepts of regression and correlation coefficients and their tests of significance should be considered as a prerequisite to learning the ratio and regression methods of estimation in sample surveys. These topics as a prerequisite to learning of sampling methods are normally taught as a part of the courses on statistical methods. However, to make the contents of the courses on sampling methods self-sufficient, the relevant topics relating to statistical methods may be considered as an integral part of the course on sampling, particularly for the students having either no training or little training in statistical methods. Most of the textbooks on sampling methods presuppose that the participants in the sampling course already have the necessary bckground in statistical methods and straightway start introducing the concept of sampling. It becomes the responsibility of the teacher to assess the background of the students and arrange the course and its level accordingly.

Sampling methods may be treated as technology for production of information on diverse fields of human activities. A professional sampling technologist must receive adequate training in all the relevant disciplines which, when used together judiciously, turn into the shape of sampling technology. As in the case of engineering, the basic disciplines being mathematics, physics and chemistry, likewise basic disciplines for sampling technology may be mathematics, theory of probability and statistical methods. In addition, a student specialising in sampling technique must acquire adequate knowledge of the area of application either through formal course work or through self-study. For example, a sampling statistician in agriculture must have adequate knowledge of the economy of agriculture and crop husbandry practices. The contents of the sampling course must, therefore, include an integrated course on the field of application. Preparation of a critical report on a sample survey project already completed may also be a part of the course contents. Such an exercise may help a student in knowing the various stages of a sample survey project.

SEQUENCE OF THE COURSE

As mentioned earlier, background in elementary algebra and calculus, and statistical methods may be considered a prerequisite to learning the subject of survey sampling. If a student has not got adequate training in these subjects, he must be advised to undergo a course containing these topics before the teacher starts the introduction of courses on sampling.

While introducing the subject of survey methodology, the first task of the teacher should be to generate interest among the students in learning it as a powerful tool for gathering information in the quickest possible time with least cost. The concept of a survey population, units of observations, character under study, parameters to be estimated etc. should be introduced with the help of live examples. The algebraic notations may be introduced only when the students have fully grasped the underlying concepts. This is the stage when merits and demerits of the census and sample survey technique as methods for collecting data should be thoroughly discussed. The communication may become more effective if the lecture is illustrated by an actual example of a census recently conducted. If the census has been conducted according to the procedure laid down and if all the arithmetical processes are correct, the census results should be free from any error. However, the same cannot be said about the results obtained from a sample in spite of the fact that all the prescribed rules have been strictly followed. This is because of the fact that the sample results are based on a fraction of the individuals constituting the population. Concept of sampling error becomes important here. The importance of sampling error as a measure of reliability of an estimate may be discussed now. Although the sampling distribution of error, unbiased estimate and confidence intervals are the topics to be generally considered as a part of the course on statistical methods, it may be useful, even at the cost of some time for repetition, to explain the underlying principles with the help of live numerical data, and electronic calculators for arithmetical calculations. The author's experience of teaching survey sampling for the last three decades to both categories of students, discipline of statistics and others, is that they understand the statistical concepts better when illustrated by actual data.

The concepts of standard error and sampling variance are of great importance in the entire theory of statistics, but these are of greater importance in developing the sample survey theory. In fact, the entire sample survey theory centers round the search for the smallest sampling variance under a given set of conditions. The students should get full insight into the utilities of standard error in presenting any sample estimate of a parameter, say mean or total. Once the students have got a good grasp of the concepts of standard error and sampling variance, the concept of non-sampling errors and statistical biases, which are not commonly understood by the non-survey statisticians, may be introduced, emphasizing their effects on the interpretation of the census and survey results. The existence of large non-sampling errors may spoil even the results of a census, which are free from sampling errors. It is one of the strongest arguments in favour of sample surveys that it is possible to control or at least minimise the effect of non-sampling errors in a well designed and well conducted sample survey, which may not be possible in a census unless enormous resources are spent for the purpose. In general, therefore, the sample estimates are subjected to both sampling and non-sampling errors. Within the framework, the efforts needed to control and balance the two errors in sample surveys are to be stressed.

The concept of sampling error is valid only under the probability sampling scheme or popularly known as random sampling. Unless the sample is randomly drawn, the calculation of samplingerror has got little meaning and statistical theory is hardly of any use to interpret the results. What is random sampling, how to draw a random sample, how to generate a population of samples from a given parental population, how the result of each random sample has got the same interpretation, etc. should be thoroughly explained with the help of numerical examples. Although the concept of random sampling that each unit of the population has got a preassigned probability of selection in the sample may be introduced in a general way, the theory of the simple random sampling, may be discussed first in greater detail as it is easier to explain the concept of unbiased estimate, the dependence of sampling variance on variations among the units and the units and the size of the sample, and the effect of the finite population. Besides deriving the algebraic expression for sample variance of the simple arithmetic mean, say $\overline{y}_n$, n being the sample size selected by simple random method without replacement, which may appeal more to the mathematically trained students, it may be interestingly shown with the help of numerical exercises as to how the sampling variance

$$V(\overline{y}_n) = \frac{N-n}{N}\,\frac{S^2}{n},$$

where N and S^2 have the usual meaning.

This is a very important formula in the theory of sampling. The development of various sampling designs can be linked with this formula.

The sampling variance of the mean depends upon S^2 and n. By increasing n, the sample size, the sampling variance can be reduced to any level. But an increase of the sample by an additional unit will result in an additional cost. Here the cost concept in sample survey may be introduced. The major advantage of using sampling method is to obtain maximum information at a given cost or alternatively to get the required information at a minimum cost. At this stage, the various sampling strategies which may help in reducing the sampling variance, particularly the factor S^2 may be introduced. In many situations, a considerable amount of supplementary information about the units of the population may be available, or can be easily collected from the records maintained by public and private agencies. Depending upon the way the information is available, the improved sampling designs may be introduced in sequence of simple random sampling. In most of the textbooks on sampling, immediately after the simple random sampling, stratified random sampling has been discussed. In the opinion of the author, immediately after simple random sampling with and without replacement, sampling schemes, selection of units with probability proportional to size (pps), and ratio and regression methods of estimation, which use supplementary information for improving the accuracy of the estimate may be taught. Courses relating to double sampling and multi-phase sampling, and even that of successive sampling may be taken as an extension of ratio and regression methods of estimation. Stratified sampling may be introduced as another way of using the supplementary information for improving the estimate. Sometimes geographical stratification is used for administrative reasons. All the topics listed above may again be discussed under stratification of the population and the theoretical implication, particularly with regard to distribution of sample size among different strata, may be highlighted.

After covering the courses relating to the sampling techniques which use supplementary information on sampling units, as well as their combinations, the cluster sampling and multi-stage sampling may be taught in sequence. Since multi-stage design is of most common use in actual sample surveys, sufficient time should be devoted to discussing various aspects, such as methods of selection of sampling units at different stages, method of estimation including ratio and regression methods, variance components, determination of the sample size and its distribution at various stages, cost function, relative efficiency, self weighting design etc.

Systematic sampling and its variations, such as linear, circular, balanced and two dimensional, may preferably be discussed as a separate topic after the students have completed the course on all aspects relating to random sampling. It may then be easy to explain its similarity to cluster sampling and stratified sampling. How important it is to avoid the systematic method of selection of the primary sampling unit in multi-stage sampling design can be better explained when the students have learnt all about the multi-stage sampling.

In a large scale complex survey, the calculation of sampling errors of various estimators becomes a time-consuming process and this is the reason that in a majority of sample survey reports, the estimates are presented without corresponding standard errors. It does not need any emphasis to say as to how difficult it is to give proper interpretations to such results. As the users are unaware of the limitation of the statistics calculated from sample surveys, they tend to utilise these figures without realising that they are subject to varying degrees of sampling and probability non-sampling errors. It may, therefore, be useful to give a course on short-cut methods of calculating the sampling errors. Many such short-cut methods have been discussed by Zarkovich (1965). In this connection, it may be useful to introduce the concept of independent interpenetrating or replicated sub-samples, as this technique is very useful for estimating sampling variances in case of complicated sampling designs. Further, this technique is also useful in measuring some types of biases and non-sampling errors.

Most of the sampling designs listed in the textbooks relate to uni-purpose or uni-character surveys. Owing to considerations of economy, operational convenience, and availability of limited qualified personnel, multi-purpose surveys have, however, become very common, particularly in many developing countries. The proposed United Nations Household Survey Capability Programme is one of the examples of multi-purpose surveys. One of the major problems concerning the multi-purpose surveys is to formulate a sampling design to estimate multiple characteristics with possibly different levels of accuracy, necessitating a search for an optimum or near optimum balance between diverse requirements, usually under cost and other operational restrictions. Although considerable work has been done on the subject and adequate experience has been gained in planning and conducting multi-purpose surveys, particularly in developing countries (Murty, 1974), it is hardly reflected in the textbooks on sampling. The teachers may have to collect the materials from statistical journals and supply survey reports and communicate the information to the students, highlighting the importance of the subject.

The last course may relate to biases and non-sampling errors in sample surveys. Although a reference to biases and non-sampling errors should always be made while initially introducing the concept of errors in sample surveys, detailed discussions on the causes of these errors and methods for minimising their effects may necessarily be conducted at the final stage of lecture courses. Fortunately, most of the textbooks on sampling have listed the causes of non-sampling errors and possible methods to minimise their effects, and these can be the basis for preparing teaching materials on this topic. However, unlike the sampling error, no systematic theory for treating all kinds of nonsampling errors has so far been developed. Each kind of non-sampling errors and biases has to be dealt with separately and this can be better supported with numerous examples of actually conducted sample surveys and censuses.

LABORATORY WORK, DEMONSTRATIONS AND SEMINARS

Survey sampling may be taught as technology for data production. Although the theory of sampling, which mainly depends upon probability and mathematics for its development, is very general, its application is broadly governed by the area of application. Therefore, the application aspects of sampling may be taught better in the statistical laboratory through practical exercises based on live data, demonstrations involving actual field work, and seminars on special types of surveys.

Lectures on a sampling design should follow a number of practical exercises. If some recent census data relating to agriculture or human population are available, that may be treated as statistical population and parameters of the population may be determined. Under various sampling designs, sample estimates may be obtained. If no census data with necessary details are available, data relating to some large scale sample survey may be used, treating it as population. From such data, depending upon the sampling design, estimate, estimate of sampling variance, confidence interval, relative efficiency of the design under realistic cost function models, the sample size needed for given precision, etc. may be worked out by each student. If in certain exercises the computation involved becomes large, the work may be suitably divided among students and results of computation may be then pooled to give the final picture.

Since the use of electronic computers and electronic calculators has now become common, the students may be encouraged to utilise these facilities in their computational work. In fact this may, incidentally, help the students in learning computer programming, now an important qualification for a statsitician.

Practical exercises may also include some empirical studies. It is now well known that there does not exist a uniformly best design in sample surveys (Godambe, 1955). Only through the empirical studies, superiority of one sampling design over the other under some conditions can at best be demonstrated.

Unfortunately, not much attention has been paid to this aspect of teaching of survey sampling. Teachers, particularly in universities, find great difficulty in gathering live census and survey data for preparation of practical exercises. To make the practical exercises effective, these should be based on national data. Realising this need, a 'Hand Book on Sampling' containing practical exercises on sampling techniques based on live data collected through sample surveys conducted in India was brought out (Singh, 1978). Similar publications are needed in other countries, where survey sampling is taught as a separate discipline.

Practical demonstration of application of sampling technique, from the initial stage of planning to the final stage of preparation of survey report should be considered as an integral part of the course. If the teaching centre itself is engaged in the conduct of sample surveys, students may be associated with the projects. They should participate in the preparation of sampling design and sampling frame, drafting of questionnaires and field instructions including concepts and definitions involved, training of field and supervisory staff, fieldwork for collection of data, scruting and editing of questionnaires received from the field, development of appropriate estimation procedure consistent with the sampling design, formulation of tabulation programme, and finally drafting of the project report. However, if the teaching centre has no such sample survey project of its own, it may formulate a live project in consultation with other teaching and research departments of the university or an agency which can finance such a project. Students under the supervision of the teachers may be made responsible for implementing it.

The demonstrations of conduct of an actual sample survey project help in establishing the firm link between sampling theory and its practical applications and bring out numerous problems faced by the survey statisticians. A student of a mathematically oriented sample survey course may not be able to imagine and appreciate the types of problems faced by a statistician responsible for planning and implementing sample survey projects unless he is brought into contact with actual work at all its stages.

The seminars should be made a regular feature of the sample survey course. The topics of common applications, but of controversial nature such as purposive sampling versus probability sampling, census versus sampling, repeat and rotation sampling, evaluation studies involving estimation of frequencies and percentages, etc. may be the subjects of discussions at the seminars. Besides students and teachers, other experienced survey statisticians employed in the government organizations should participate in such seminars.

TEACHER AND INSTITUTION

It may not be out of the way, at this stage, to say a few words about the qualifications of the teachers and teaching institutions. No matter how fine the materials and contents of the course may be, ultimately it is the quality of human resources, which decides success of a programme. We are well aware of the necessity of high standards of training in both theory and practice. Since sampling is basically an applied subject, a balance between the theory and its practice has to be maintained while organising the course. The effectiveness of the teaching programme will ultimately depend upon the qualities of teachers and the taught.

The teacher of such a course should have training not only in mathematical aspects of the theory of sampling, but should also have adequate experience in its application for collection of data and their interpretation. Unfortuantely, the teachers in most of the universities which are the main centres for teaching of survey sampling, have hardly any orientation in the application aspect. As a result, the training remains lop-sided and when a student after successfully completing such a course takes an appointment in a private or public undertaking where his main job is to design and conduct sample surveys he finds himself completely unfit and loses confidence in his ability to do the job. That is why a paradoxical situation has arisen in many countries that unemployed statisical personnel and vacant positions exist side by side.

The universities and other teaching centres have, therefore, to orient their teaching programmes to meet the pressing needs of the country. Ideally, the best institutions for teaching an applied subject like survey sampling are those which are simultaneously responsible for an integrated programme of education, research and extension. Conceptually, extension may mean here the use of training and research to solve the practical problems faced by the country. This system will automatically produce teachers who are fully acquainted with the latest development of the subject through research, and have realisation of the relevance of the course through practical applications. The author discussed this particular aspect in a more general way in his presidential address at the Indian Science Congress (Singh, 1976).

The question is now whether the universities are capable of developing such a system or alternatively, whether new institutions with specific objectives in view can be created. Happily, some such institutions, to name a few, Survey Research Centre, Institute of Social Research, Univeristy of Michigan; International Statistical Programme Center, US Bureau of the Census, Washington D.C.; Indian Statistical Institute, Calcutta; Indian Agricultural Statistics Research Institute, New Delhi etc., with an integrated programme of teaching, research and extension of survey techniques already exist. The requirement of survey statisticians is indeed great. Many more such centres need to be established. Some of the universities with strong departments of mathematical statistics and other necessary infrastructures may orient their teaching and research programme to meet the requirement of trained survey statisticians. The International Statistical Institute with the active cooperation of the International Association of Survey Statisticians, and with the financial aid from international agencies, may consider the establishment of some centres for training and research in survey sampling to meet the needs of developing countries. The Institute has already shown the way by organizing the World Fertility Survey on a world-wide basis.

CONCLUDING REMARKS

Sampling technique may be considered as technology for information production needed for planning socio-economic development. In teaching the sampling course, a balance between sampling theory and its application is required to be maintained. This can be achieved only if a course on design follows practical exercises based on live data. Demonstrations of conduct of actual surveys by involving students from the initial stage of planning to the final stage of reporting the results may be considered as an integral part of the course. A proper sequence in planning the course may be helpful from the communication point of view. Adequate emphasis on the role of sampling and non-sampling errors in interpreting the survey results may be laid. To maximise the utilities of training and research in sampling techniques, there is a need of developing an integrated programme of teaching, research and extension. To meet the increasing demand of trained survey statisticians, particularly in the developing countries, the teaching programme in universities may need orientation. There may be justification for establishing regional centres for training and research in sampling in developing countries with the aid of international financing agencies. The International Statistical Institute may take a lead in the matter.

References

[1] Cochran, W.G. (1953,1963) Sampling Techniques. John Wiley & Sons, New York.

[2] Deming, W.E. (1960) Sample Design in Business Research. John Wiley & Sons, New York.

[3] Des Raj (1968) Sampling Theory. McGraw Hill Book Co., New York.

[4] Godambe, V.P. (1955) A unified theory of sampling from finite population. Journal of the Royal Statistical Society, (B), 17, 269-278.

[5] Hansen, M.H., Hurwitz, W.N. and Madow, W.G. (1953) Sample Survey Methods and Theory, Volumes I and II. John Wiley & Sons, New York.

[6] Kish, L. (1967) Survey Sampling. John Wiley & Sons, New York.

[7] Murty, M.N. (1967) Sampling Theory and Methods. Statistical Publishing Society, Calcutta.

[8] Murty, M.N. (1974) Evaluation of Multi-subject Sample Survey Systems. *International Statistical Review*, 45, 175-192.

[9] Singh, D. (1976) *Economic Planning of Rural India, Statistics, and Statisticians*; Presidential Address, Statistics Section, Indian Science Congress, 63rd Session, Indian Science Congress Association.

[10] Singh, D., Singh, P. and Pranesh Kumar (1978) *Handbook on Sampling*. Indian Agricultural Statistics Research Institute, New Delhi.

[11] Sukhatme, P.V. (1953) *Sampling Theory of Surveys with Applications*. Iowa State College Press, Ames, Iowa, second edition with Sukhatme, B.V. (1970).

[12] United Nations (1947-1951) Reports of the Sub-commission on Statistical Sampling; E/CN 3/37, 52, 83, 140, New York.

[13] Yates, F. (1949,1960) *Sampling Methods for Census and Surveys*. Charles Griffin & Co., London.

[14] Zarkovich, S.S. (1965) *Sampling Methods and Censuses*, FAO, Rome.

TEACHING OF SURVEY SAMPLING

P K Bose, Calcutta University

Statistical thinking is a form of logical thinking and is a part of our daily lives. When we say something or somebody is typical we are thinking in terms of statistical averages and departure from type is statistical variation. When we generalise from a few cases to a very large number, we employ sampling processes. A conclusion that two things always go together involves a pattern of thinking found in statistical correlation. Statistical thinking is not foreign to everyday thinking but is a scientific form of it. Statistics is a fundamental activity of mankind.

At present we are noticing two important trends in statistics which need some rethinking in teaching statistics in the Universities/Colleges and other institutions. One is its massive impact on society and the other is the extensive growth of statistical theory.

The merging of these two trends into a single stream makes new demands on our statistical personnel. Neither the isolated theorist nor the practitioner is able to keep abreast of the current. The specialist in mathematical statistics must acquaint himself with statistical problems and equally working statisticians must familiarize themselves with the logic of statistical theory.

It is necessary to bring students into awareness of and harmony with the statistical content of society. This content is extensive. Such large scale activities should be integrated with academic concepts such as probability, sampling, estimation and testing of hypotheses.

It is felt that the tools of statistics should be brought out and sharpened only after the need for them has been realised.

Reasoning from sample to universe is called statistical induction. Inductive reasoning involves observing individual cases and arguing from them to a general principle. In statistical induction we make inferences concerning the characteristics of a large group, sometimes of an infinitely large group, from the characteristics of a specifically chosen, usually small, part of the group that we observe. We draw conclusions from the smaller group and apply these conclusions to a larger mass.

In statistical induction we are able to add to our generalisations a statement as to their precision.

It involves the following stages

1. Define the trait to be measured.
2. Measure this trait in a sample.
3. Apply the measure to the universe from which the sample is drawn.
4. Estimate the error involved in this application.

Basic sampling technique deals with the four stages enumerated above.

For teaching survey sampling the first course should be statistical reasoning, it should be followed by major statistical tools and the third course should be sampling methods.

Sampling plays a vital role in research design involving human populations, public health, business administration, public administration, commerce and industry and various disciplines of social, biological and physical sciences.

Excepting routine teaching on sampling in various universities and colleges, a very few institutions organise training programmes on survey sampling. In the Indian Statistical Institute we carried out experiments in statistical survey sampling for over a decade which has laid the foundation of survey sampling on a large scale in the world.

Sample enquiries fall into two broad categories. One is sample survey on an extensive scale covering a wide area in which the object is to obtain reliable estimates of the acreage, rate of yield per acre and total production of food and fibre crop and of economic or demographic factors relating to indebtedness, unemployment, distribution etc. The second type of work is an enquiry which is more localised and related to cost and level of living, housing, consumption of food etc.

The difference between the two is great so far as quantity is concerned and this brings in practice a change in quality. As Professor P.C. Mahalanobis suggested, large scale sample surveys may be appropriately called statistical engineering. Teaching of survey sampling should be concerned more with statistical engineering rather than with the pure theory of sampling.

In preparing the design of each sample survey the three most important considerations are time, cost and the human agency - all of course in relation to the permissible margin of error. I have already indicated that the teaching of survey sampling should consist of three sections, namely fundamentals of statistical reasoning, statistical tools and sampling methods. I am not discussing the first two; they are readily available in most text books on statistics.

Regarding the third I have already discussed the essential difference in concepts between a small sample enquiry and large scale sample surveys. Sampling theory as applied to small or localised enquiries is also elaborately dealt with in many text books. Teaching is carried out mostly on the basis of such books.

Unfortunately, very little attention is given to experiments to be carried out in the laboratories; unless the students carry out some experiments themselves, the fundamentals of sampling will never be clear.

I may give a few illustrations.

Experiment one

Conduct opinion poll sampling of student body; construct questionnaire, design sample, discuss interview technique, tabulate results, make inferences about population.

If the above experiment is carried out in the laboratory the students will enjoy this experience, and will ask many penetrating questions. This may open opportunities for sound instruction.

Experiment two

Extend sampling experience by use of coins or dice or cards. Record result of each toss or draw.

Verify probability laws empirically.

Experiment three

Use bowl containing equal number of beads of two colours or use table of random numbers with equal numbers of even and odd digits. Draw many samples of ten and record the number of successes.

Emphasise sampling variation and observe improved reliability when samples are combined into large samples.

Method of teaching statistics, and more especially sampling, should be based on experiments which combine observations, calculations and inference.

In teaching sampling theory emphasis should be laid on this aspect.

The second important aspect which is overlooked while teaching is the quality problem of statistical data. No method of collecting and processing data is safe from errors. Some ideas about errors that statisticians are expected to study and evaluate should be given. Errors in the broad sense of the word may be classified in three groups.

A. Errors resulting from inadequate preparations.
B. Errors committed in the stage of data collecting.
C. Processing errors.

Necessary precautions should be taken for reducing these components of errors. Large scale sample surveys which form the major part of survey sampling have a number of distinguishing features, namely the concepts of optimum design of surveys, pilot surveys and inter-penetrating network of sub-samples and the control of non-sampling errors. In survey sampling the magnitude of non-sampling error which is due to various causes is more than the sampling errors. A detailed discussion of these topics should form part of a teaching programme.

The pioneering work of P.C. Mahalanobis in the field of large scale sample surveys is widely recognised. He demonstrated the possibility of using large scale surveys to collect information quickly and economically with sufficient accuracy.

COST FUNCTION

Large scale surveys are necessarily expensive and considerations of cost are therefore of great importance. Total cost of operations depends on the size of the grids and their total number. Working within a fixed budget this means once the size of grids is settled, their total number must also be determined by the budget limit.

Professor Mahalanobis obtained the total cost function in his study on sample census of Jute as

$$\phi = c_o + c_1 y_i - c_2 y_i^2 + c_3 x_i \; y_i$$

where x = size of sample unit and y = density of sample unit.

VARIANCE FUNCTION

The precision of the final result is important. The variance of grids or sample units decreases as the size of the grid is increased.

Professor Mahalanobis obtained the variance function as

$$v = \frac{\Sigma A_i \; \alpha \; p_i \; q_i}{y_i \; (x_i)^{\beta}}$$

where A_i = Area, p_i = proportion of land under crop.

OPTIMUM DESIGN

It is now possible in terms of the cost and variance functions to state the conditions for an optimum or most economical design for sampling. The aim may be stated in either of the following two alternative forms:

a) to determine the size and distribution of sample units in such a way that the variance of the final estimate is a minimum when the total cost of the survey is fixed;

or

b) to determine the size and distribution of sample units so as to reduce the variance of the final estimate to any desired level at a minimum cost.

Professor Mahalanobis obtained the most economic design for sampling with the help of an empirical knowledge of variance and cost functions.

SEQUENTIAL DEVELOPMENT

In large scale work, the survey has often to be continued from year to year or at suitable intervals. It is possible to adopt an exploratory method in which a sample survey is carried out on a very small scale with the primary object of collecting basic information required for preparing an efficient design for later surveys.

In the exploratory stage the object is to gather as quickly as possible some rough idea of the variance function, cost of operations from one region to another type and nature of human agency likely to be suitable for field work. As basic information begins to be gathered the design of the survey steadily improves and the cost per unit rapidly decreases in subsequent surveys.

INTER-PENETRATING SUB-SAMPLES

One of the chief objects of the sample survey is to test the accuracy of results attained in practice by a field trial on a large scale. The information is collected in the form of two independent half-samples, each of which consists of nearly half the total number of sample units. In each half-sample the sample units are located at random over the whole area and the field operations for each half-sample are carried out by different sets of investigators. The difference between the results obtained from the two half-samples will furnish a good idea of the margin of error inherent in the present method.

The teaching of survey sampling units should include mainly the following.

A. Introductory courses on statistical reasoning and statistical tools.

B. Fundamentals of survey sampling.

C. Special problems etc. and techniques of large scale sample surveys.

Topics related to A and B are well known. They have been discussed in different platforms and good text books are also available on them. Regarding C it is still uncharted. A few items covering the same are given below.

(a) Terminology and classification of sample surveys.

(b) Cost function; variance function; optimum solutions.

(c) Sequential development technique.

(d) Planning of surveys.

(e) Classification of errors.

(f) Controls at the stage of statistical processing.

References

[1] Mahalanobis, P.C. Experiments in Statistical Sampling. Asis Publishing House, Calcutta.

[2] Mahalanobis, P.C. Sample Census of Area under Jute in Bengal 1940. Statistical Publishing Society, Calcutta.

[3] Mahalanobis, P.C. (1944) Large scale sample survey. Philosophical Transactions of the Royal Society, London Series, 231.

[4] Sukhatme, P.V. (1947) Problem of plot size in sample surveys, JASA.

[5] Snedecor, G.W. (1948) A proposed basic course in statistics, JASA.

[6] Seng, Y.P. (1951) Historical surveys of the development of sampling theories and practice, JRSS (A).

[7] Zarcovich (1956) History of sampling in Russia, JRSS (A).

TOPIC 13: TEACHING DESIGN AND ANALYSIS OF EXPERIMENTS

ORGANISER: DR D A PREECE

THE DESIGN AND ANALYSIS OF EXPERIMENTS - WHAT IS IT ABOUT?

D A Preece, Rothamsted Experimental Station

ABSTRACT

Basic definitions are given. Then the subject is divided into six "components", which are discussed in turn:-

I: Planning, design and layout.
II: Management of the experiment or experiments.
III: Data recording.
IV: Scrutiny and editing of the data.
V: Computational analysis.
VI: Interpreting and reporting the results.

Over-theoretical, over-mathematical teaching of the subject is criticised. The importance of practical considerations is stressed.

Key Words:
Analysis of variance; Balance; Blocking; Checking copying and calculations; Confounding; Covariance analysis; Data recording; Data scrutiny; Experimental error; Factorial designs; Null hypotheses; Randomisation; Repeated measures; Replication; Report-writing; Significance; Split-plot designs; Transformations.

INTRODUCTION

Within the subject of this paper, an 'experiment' is 'comparative' (Anscombe, 1948; Cox, 1958, Section 1.1; Finney, 1960, Section 1.2): it compares different "treatments", which might be different varieties of wheat, different diets for pigs, different drugs for insomniacs, or different acid concentrations in an industrial process. Many comparative experiments are "factorial" (Yates, 1937; Cox, 1958, Chapter 6), permitting simultaneous study of different "factors", each at different "levels". Thus in a fruit-tree experiment, three pruning methods (factor A, with 3 levels) might be tried in conjunction with four sprays (factor B, with 4 levels); in this "3x4 factorial experiment", the twelve combinations of pruning method and spray are the 12 "treatment combinations". A factor may be "qualitative" (e.g. different colours in a psychological experiment) or "quantitative" (e.g. different doses of Vitamin C).

Each experiment comprises a set of "experimental units" (often called "plots", and sometimes "runs"), with different treatments or treatment combinations applied to different experimental units. These may be, for example, pens of animals, individual animals or plants, individual leaves on growing plants.

or (in an inoculation experiment on plants) half-leaves. Commonly, treatments are "replicated", i.e. there is more than one plot per treatment. However, factorial experiments include "single replicate experiments" (one plot per treatment combination) and "fractional replicates" (not all treatment combinations present).

The comparison between treatments, between factor levels, and between treatment combinations are based on data recorded after the treatments have had a chance to affect the experimental material. These data are one or more "variates" (often called "variables" - although this term is sometimes used for "factors"), each variate consisting of one value for each plot (except possibly for "missing values"). Variates may be "continuous" (e.g. weights, heights, temperatures, pressures) or "discrete" (counts, scores); they may be "primary" (obtained by direct reading of a scale, by counting, or by scoring) or "derived" (obtained from one or more other variates by arithmetic calculation). In statistical terminology, "analysis of an experiment" is elliptic for the analysis of a variate or variates, possibly in conjunction with other data pertaining to the experiment. For a factorial experiment with quantitative factors, the analysis of a variate may consist of investigating a "response surface"; for present purposes, experiments designed to study response surfaces are regarded as comparative.

"Design" is more difficult to define than "analysis", not least as it means both "designing" (part of "planning") and "a specific design". Finney (1955, Section 1.2; 1960, Section 1.3) defined the "design" of an experiment to mean

(i) The set of treatments selected for comparison;
(ii) The specification of the plots;
(iii) The rules for allocating the treatments to the plots;
(iv) The specification of the measurements or other records to be made on each plot.

A fifth component should perhaps be listed as being a legitimate concern of "designers", whether or not it is strictly a part of design:

(v) Specification of details of the management of the experiment.

But the "design and analysis of experiments" deals not only with individual experiments, but also with series of experiments, indeed with whole experimental programmes. Within a series, there may be experiments in different locations, or on different occasions (perhaps in different years), or both (John and Quenouille, 1977, Chapter 12).

The need for a statistical subject of "design and analysis of experiments" arises from variability inherent in the experimental material, environment, and management. This variability leads to uncontrolled – indeed uncontrollable – variability in the observed variate values, i.e. to "experimental error", where "error" does not mean "mistake". A variate affected differentially by the treatments thus includes variability from one treatment to another, plus the "error" variability that would have been present even if a single treatment (or none) had been applied throughout. The experimental error can be regarded as "background noise" such as the "static" heard when radio reception is poor. Differences in effect between treatments can then be likened to a broadcast message; if the message is spoken quietly (treatment differences small), it may be drowned out by the static.

COMPONENTS OF THE SUBJECT

The main principles, tasks and skills of our subject can perhaps be summarised in the following "Top Twelve" headings, which I have tried to arrange in a logical order:

(1) The basic principles of statistical inference.
(2) The ability to select treatments, experimental material, experimental units, and variates that will permit "useful" results to be obtained.
(3) The ability to recognise relevant sources of variability, and to make preliminary assessments of their importance and relative magnitudes.
(4) The basic principles of "good" design.
(5) Mathematical methods for constructing designs and obtaining analyses, particularly analyses of variance and covariance.
(6) Trained personnel for setting up and managing experiments.
(7) Proper instruments (balances, etc.) for obtaining accurate, precise data values, and trained personnel to use the equipment.
(8) Training in data transcription (with checking).
(9) Ability to recognise and interpret patterns and anomalies in data.
(10) Understanding of the assumptions underlying standard statistical analyses, and the ability to recognise when these assumptions are seriously violated in practice.
(11) Facility in spotting mistakes in analyses.
(12) Skill in reporting conclusions in clear, simple prose, with appropriate tabulations and graphs.

In the teaching of the design and analysis of experiments – whether in textbooks or lecture-rooms, whether in syllabuses or examinations – some of the above topics have received much less time and attention than they deserve. The emphasis has often been on the theoretical and the mathematical, to the detriment of the applied and the practical. Derivable formulae, provable theorems and repeatable mathematical procedures have taken precedence over studies of the environments in which experiments are done, over getting to proper "feel" for experimental data, over thinking about the validity of assumptions and of results, and over report writing. The emphasis given to different parts of the subject must, of course, depend on the audiences, but insufficient attention to the Real World of real experiments transforms the subject from statistics to mathematics, from science to ritual (Preece, 1982b).

So the words "design and analysis" of themselves provide a very deficient picture of the content of our subject. To avoid both the over-simple dichotomy between "design" and "analysis" and the over-elaborate set of twelve headings above, I now break up the subject into six components, each demonstrably of statistical concern, that may be labelled roughly and simply as follows:-

I Planning, design and layout;
II Management;
III Data recording;
IV Scrutiny and editing of the data;
V Computational analysis;
VI Interpreting and reporting the reults.

Elsewhere (Preece, 1982a) I did not include the present component II; however, as bad management of an experiment can introduce major variability and biases, I now believe that management merits specific mention. Other lists can of course be produced, with different amounts of detail and different emphases; the nineteen "Main Features of a Clinical Trial" given by Pocock (1982, p.3) are a good example.

COMPONENT I: Planning, Design and Layout

In the vocabulary of Jeffers (1978), this first of my six components includes

(a) Stating the objectives [of an experiment];
(b) Defining the population about which the inferences are to be made;
(c) Selection of experimental treatments;
(d) [Choosing the] Plot shape and size.

The choices (c) and (d) cannot be made separately from the choices of treatments to be included and of number of plots; all these choices, and the allocation of treatments to plots, are to be made in accord with the three basic principles (or "three R's") of Fisherian experimental design: Replication, Randomisation and Blocking. The three were discussed in the first edition of "Statistical Methods for Research Workers" by R.A. Fisher (1925); their explicit formulation can be taken to mark the start of the statisical subject of the "Design and Analysis of Experiments". [Writing for agricultural research workers, Little and Hills (1978, p.9) listed a different "three R's": Replicate, Randomize, and Request help! Das and Giri (1979, p.7) gave the Fisherian three as

Randomization, Replication and Error Control, with "Error" again meaning "experimental error".]

Replication of a treatment involves the treatment being applied to more than one plot; the number of plots carrying the treatment is the number of "replications" or "replicates" of the treatment. Confusingly, the word "replication" is also used for a set of plots within which each treatment occurs exactly once. Replication enables estimates of experiments error to be made.

Randomization (Cox, 1958, Chapter 5; Finney, 1960, Section 1.6) is a procedure for allocating the treatments to the plots. The allocation adopted is chosen at random from a set of possible allocations, the set being such that the experiment shall provide unbiased estimates of treatment differences and of experimental error. In many experiments, randomization gives all the plots an equal chance of receiving any particular treatment or treatment combination; there will not however be equality in a properly randomized experiment with twice as many plots for a control treatment as for any other treatment. Randomization avoids the biases that can arise from subjective allocation of treatments to plots.

Blocking is a grouping of the experimental units into "blocks" (often all containing the same number of plots) with the aim of reducing the experimental error or "background noise" in variates to be studied. The grouping is done with the hope of capturing any large components of the noise as "between blocks" variability which can then be filtered out in the statistical analysis: plots likely to give the most similar results in the absence of treatment effects are, as far as possible, grouped together within blocks. Perhaps the commonest experimental design is the "randomized complete block design" (often just "randomized block design"), in which each block contains exactly one plot for each treatment, the allocation of treatments being randomized independently within each block; this design is used for many agricultural field experiments, often with each block a compact group of adjacent plots. Some experiments, notably on fruit-trees in orchards, take the idea of blocking further, in having a "row-and-column design" with two overlaid systems of blocks, namely "rows" and "columns", at right angles to each other.

Adherence to the Fisherian "three R's" helps to satisfy the "Requirements for a Good Experiment", which Cox (1958, Section 1.2) gave as

(A) Absence of systematic error;

(B) Precision (suitably small random errors of estimation);

(C) A wide range of validity for the conclusions that are to be drawn;

(D) Simplicity;

(E) A proper statistical analysis to be available which, without artificial assumptions, permits calculation of the uncertainty in the estimates of treatment differences.

Because of the need to choose suitable experimental material and an appropriate design for that material, it is often desirable for statisticians involved to visit the location of an experiment, to join the experimenters in examining the material, and to have discussions not only with the organisers but also with supervisory and experimental-management staff.

Planning discussions about (a), (b) and (c) must often cover the possible inclusion of a control treatment. Sometimes a strictly "untreated" control may seem appropriate, sometimes - in medical vocabulary - a "placebo".

Basic ideas about factorial experiments were set out by Yates (1937) and rehearsed by Cox (1958, Chapter 6). Fisher (1926) pointed out that an experiment studying several factors simultaneously can have various advantages over a succession of experiments each on one factor only; as paraphrased by Yates (1964, p.312), the advantages are:-

(1.) the plots are used several times over to determine the average effects of the different factors;

(2.) only by factorial design can any information be obtained on how responses to one factor are affected by another (i.e. on interactions);

(3.) factorial experiments provide a wider inductive basis for conclusions on the effects of the factors.

Some factorial experiments have "split-plot" designs, the simplest of which has two factors, A and B, with the levels of A allocated to "main plots" or "whole plots", each subdivided into as many "sub-plots" as B has levels; the levels of B are assigned at random to the sub-plots within each whole plot, a separate randomisation being used for each whole plot. Because of the two sizes of plot, and the different randomisations for A and B, the analysis of variance for the design has two error terms - "whole-plot error" and "sub-plot error" - not just a single "(experimental) error". There are three possible reasons for using a split-plot design (Finney, 1960, Section 5.13, p.85):-

(α) Practical considerations (e.g. size of machinery) may make it undesirable or impossible to apply the levels of some factor(s) to small plots, whereas these levels may be feasible on large subdivided plots.

(β) an experimenter may wish to add some further factor(s) into an existing experiment, and this may be possible by putting the new factor(s) "on sub-plots" within the existing plots.

(γ) Often, error variability between whole plots is greater than that between sub-plots; then, the effects of sub-plot factors, and the interactions between sub-plot and whole-plot factors, are estimated more precisely than if the sub-plots were not grouped into whole plots, whereas there is a loss of precision on main-plot factors.

The last of these reasons seems to be of little importance in practice. Despite reports of mis-use of split-plot designs (e.g. Mead, 1982), plot-splitting is very important in the design and analysis of experiments. The omission of split-plot designs from an admittedly mathematical "systematic treatise" on factorial designs (Raktoe, Hedayat and Federer, 1981) is unsatisfactory and bizarre.

Some factorial experiments have fewer plots per block than there are treatment combinations. Important effects of the factors are, however, still estimable if suitable subsets of the treatment combinations are allocated to the blocks, the subsets being obtainable from simple mathematical relationships that define a "confounding scheme". Then the design may be said to be "confounded"; more strictly, certain effects (usually interactions between as many factors as possible) are "confounded with blocks", totally or partially (Yates, 1937; Cox, 1958, Section 12.3). Interactions that have been confounded are usually ones in which the experimenter has little interest, or which the experimenter believes to be negligible.

Some non-factorial designs have fewer plots per block than there are treatments. Such "incomplete block designs" include the "balanced incomplete block designs" (Cox, 1958, Section 11.2; Finney, 1960, Sections 6.2 to 6.6) that have been of so much interest to combinatorialists.

COMPONENT II: Management

The administration, supervision and management of an experiment need systematic care if reliable results are to be obtained. If the plots are pieces of land, these must be measured up and planted properly. If the treatments are different rates or doses, these must be determined accurately. Steps must be taken to try to ensure that each plot receives the intended treatment or treatment combination; for field and orchard experiments, this requires a plan of the physical layout, whereas various options are available for clinical trials (Pocock, 1982, p.4).

Management of a field experiment continues to be statistically important after the treatments have been applied. Extra variability can be introduced into the results by differential management of the plots, e.g. by having some plots weeded more carefully or frequently than others in a field experiment where treatment differences are not intended to include weeding differences. Such circumstances may require different blocks to be weeded by different people, so that unwanted "between labourers" differences are assimilated in "between blocks" differences. Likewise, if an experiment has to be irrigated throughout, but there is insufficient equipment to irrigate it all at once, the irrigation should be "by blocks". Also, such operations as scoring for disease - and usually harvesting itself - should be "by blocks" (Dyke, 1974, p.43).

COMPONENT III: Data Recording

The need for suitable balances, etc., for obtaining accurate, precise data values has already been mentioned. When weights, lengths, and other continuous variates are to be recorded, a "degree of precision of recording" should be specified and adhered to, e.g. "to the nearest 5gm", "to the nearest 25cm", etc. A handy rule-of-thumb specifies a variate as having an acceptable degree of precision of recording if the "grouping interval" (5gm, 25cm, etc.) is less than *one-quarter* of the expected standard error per plot (Cochran and Cox, 1957, Section 3.33, p.60). Sometimes, up to *one-half* of the expected standard error per plot is considered acceptable (e.g. Dyke, 1974, Appendix A, pp.162-164).

Data should be recorded on prepared data sheets having proper headings. Copying of data introduces errors and should be avoided if possible; unavoidable Copying should be checked by "calling over" (Dyke, 1974, p.49), and good practice requires copies to be signed by the two people involved in the calling over.

In field experimentation and industrial experimentation, some variates are obtained from sampling within the plots (see, e.g. Gomes and Gomes, 1976, Chapter 13). A pre-determined sampling procedure must then be adhered to for each plot, so that the recorder's selection is not influenced by what is found in a plot.

Variates consisting of scores need special care, as scoring can be very subjective. Scoring schemes need to be well worked out in advance, and scorers need training and practice. Scoring should usually be "by blocks" (Dyke, 1974, p.43).

Data recorders themselves sometimes calculate derived variates. Such calculations should be checked, and everyone involved should have been taught that the *worst* way to check calculations is merely to repeat them.

COMPONENT IV: Scrutiny and Editing of Data

Science is not well served when experimental data are merely "processed" in a standard computer analysis, without anyone first "sniffing over" the data to note any salient patterns or anomalies. All too many experimenters do not adopt a critical attitude towards their data and, in this respect, cease to be scientists. Anomalous values (e.g. values much larger or smaller than might be expected, values not recorded to the chosen degree of precision, etc.) inevitably occur, and some can readily be found to result from erroneous copying or flawed experimentation; also, an unexpected value could well be the most important result from an experiment.

Skill in "sniffing over" data is an important part of "Statisticianship", which is to statistics what musicianship is to music; a person without musicianship (or statisticianship?) hardly deserves to be called a "musician" (or "statistician"?). Data-sniffing skill is gained through experience rather than learnt (as are many of a musician's skills), but opportunities for gaining some of the experience can easily be provided when the design and analysis of experiments is being taught. Some statisticians claim to handle far too much data for them to have time to look through most of the data before analysis; however, a statistician with a well-developed data-reading skill can pick out important anomalous values very quickly indeed, with the ease with which an orchestral conductor, with his score-reading skill, can pinpoint misprints in a full orchestral score. Similarly, skill in detecting heterogeneity of error can be highly developed.

Editing of data clearly involves correction of known mistakes, but what is to be done about unexplained outliers and other doubtful observations? Hard and fast rules cannot be provided for this, and judgement must often be used. Discussion between experimenter and statistician may be needed or desirable.

COMPONENT V: Analysis

Writing of the experimental designs considered in his text-book, Lewis (1968, p.11) said that:

> All assume measurement on an ordinal or interval scale
> and so permit an analysis of variance.

Even apart from the unsatisfactory use of "assume", this statement is dangerous, in that it might lead the unwary and unthinking - and the educational researcher for whom the book was intended - to suppose or believe that any continuous variate, or any variate consisting of scores or counts, can reasonably be analysed in an analysis of variance. Sadly, many researchers do indeed seem to have such a belief, and to perceive analysis of variance as a "black box" for converting almost any variate into publishable tabulations and into statements about "significance". Such uncritical use of analysis of variance can lead to publication of tables such as the following (ICRISAT, 1981, p.61, part of Table 8), where I have substituted code letters for treatment names:-

Treatment	Downy mildew incidence (%) at		
	26 days	47 days	73 days
A	0	0	0.6
B	0	0	0.9
C	0	0.7	1.5
D	0	1.9	3.5
Untreated	63.9	64.3	65.0
SE±[a]	(0.72)	(1.20)	(1.63)

[a] Standard error of mean, based on arcsin transformation of the percentage values.

Here, each standard error is presumably intended to apply to each of the five means above it, and seems to have been calculated from all plots of the experiment. But this is absurd as, at 26 days, there was no mildew on any of the treated plots, and so no variability of incidence on these plots. (The standard errors are also glaringly wrong in having seemingly been obtained by illegitimate back-transformation from standard errors calculated for means of angular-transformed percentages. And the use of the "angular transformation" - mentioned later in this section - is unlikely to have been appropriate even where the means are not zero).

R.A. Fisher's well-known statement (Fisher, 1935, Section 8) that

> Every experiment may be said to exist in order to give the facts a chance of disproving the null hypothesis.

is not above criticism. A factorial experiment can be said to have several null hypotheses (for "main effects" of individual factors, and for "interactions" between factors), not just one. But more fundamentally, as Yates (1964, p.320) pointed out:-

> In many experiments ... it is known that the null hypothesis customarily tested, i.e. that the treatments produce no effects, is certainly untrue; such experiments are in fact undertaken with the different purpose of assessing the magnitude of the effects. Fisher himself was of course well aware of this ...

Estimated means and standard errors are thus the main if not sole aims of many analyses of experimental results. Even when significance tests may seem to be called for, we must remember a point made by Cox (1977, p.61):-

> statistical significance is quite different from scientific significance ... therefore estimation, at least roughly, of the magnitude of effects is in general essential regardless of whether statistically significant departure from the null hypothesis is achieved.

A continuing need to stress the true role of significance tests in experimental work has recently been recognised by authors who have contributed the following to the medical literature: "The inadequacy of statistical significance" (Lutz and Nimmo, 1977); "Significant or important?" (Wade and Waterhouse, 1977); "Significance, importance and equality - three basic concepts in the analysis of a difference" (Taube, 1980).

Overemphasis on significance testing or hypothesis testing has been carried to extremes by workers who seem to believe that "multiple comparison" tests (see, e.g., O'Neill and Wetherill, 1971) must always be applied to treatment means from an experiment. "Duncan's multiple range test" is particularly conspicuous in many journals of agricultural research; indeed Martynov and Krupnov (1980) reported that, of one hundred publications in the *Agronomy Journal* for 1973, sixty-eight used the Duncan test. Particularly disheartening are two common "glaring abuses" that were roundly condemned by Chew (1976): the use of a multiple range test in disregard of a natural ordering to the treatments and in disregard of a factorial structure to the treatments.

Some theoretical background to the analysis of variance must of course be taught to students of the design and analysis of experiments. But much of the supporting mathematical and conceptual framework in the literature is unnecessary and indeed harmful. For example, as Pearce (1982) wrote in a book-review:-

> one experimenter may be studying four inhibitors at two temperatures, while another has four diets, which he tries out on rats on two occasions. Why regard them together as examples of "two-way classifications" and then go on for pages about model I and model II, fixed and random effects? The first investigator is interested in both inhibitors and temperatures and in the way they interact, the other only in diets, the occasions merely providing confirmation. Their logical problems are unrelated.

A student, faced with data from either of these experiments, ought to be able to use a desk- or pocket-calculator to produce a correct analysis, without conscious reference to an algebraic formulation.

Likewise, a student faced with data from a simple split-plot experiment ought to be able to produce a standard analysis of variance without recourse to cumbersome mathematics. Except for the calculation of some of the standard errors, the analysis of variance for a split-plot design is simple (Preece, 1982a, Section 6.8), and not the perilous matter that it has sometimes been presented as.

Some experiments involve measuring the same attribute on several successive occasions on each plot, the occasions being the same for all plots; each plot receives its treatment just once, before the first measurement, or receives the same treatment repeatedly or continually. Such an experiment is often said to be a "repeated measurements experiment" or to have a "repeated-measures design", although "successive" would be much better than "repeated". Sometimes the successive occasions for each plot have been likened to sub-plots; indeed Little and Hills (1978, p.137) horrifyingly said that repeated-measurements data "are most properly analyzed as the split-plot design", and Aitkin (1981) that repeated-measurements designs "are called 'split-plot' designs". These statements notwithstanding, the two situations are formally very different: the split-plot design has randomisation within each whole plot, whereas the successive occasions cannot be randomised and have an inherent ordering unparalleled in the split-plot design. In some circumstances, a "split-plot" analysis may perhaps turn out to be a satisfactory approximation for a repeated-measurements analysis. However, the important paper by Rowell and Walters (1976) considered five diverse repeated-measurements experiments for all of which a "split-plot" analysis was manifestly invalid, and suggested that a simple alternative analysis should always be used for "repeated observations"; the alternative consists of separate analyses of different "contrasts over time", each contrast having been evaluated for each plot.

Preliminary scrutiny of a variate may suggest that it does not satisfy the assumptions lying behind standard analyses, but that the assumptions might be satisfied after the variate has been "transformed". Several "transformations" are commonly used, including the "logarithmic", "square root" and "angular". A log transformation, for example, simply involves replacing each variate value x by log x. The angular transformation, useful for *some* variates consisting of percentages or proportions, involves replacing each proportion p by arcsin $\sqrt{p}$, i.e. by $\frac{1}{2}$ arc-cos (1-2p); this was presumably the transformation used in the ICRISAT analysis mentioned above. Rules-of-thumb for the choice of transformations have their place in our subject (see, e.g., Goulden, 1952, p.98), but should be used critically.

"Covariance analysis" is another important topic within our subject. At its simplest, covariance analysis involves two variates (each with one value per plot), which may be denoted as y, the variate of primary interest, and x, the "covariate" consisting of "concomitant observations" unaffected by the treatments (Cox, 1958, Chapter 4). Often, the x-values are obtained from the plots before the treatments are applied. the supplementary information from x is used with the hope of reducing the experimental error in y, and so as to "adjust" the y treatment means for initial plot differences in x. The analysis is a combination of linear regression analysis and of a standard analysis of variance of y. Covariance analysis can also be used, with care, with an x-variate that *is* affected by the treatments (Cochran and Cox, 1957, Section 3.88).

Combining results from different experiments within a series is yet another important analysis topic, but is too complex to discuss here. Sadly, standard textbooks give it little attention, except notably those of Cochran and Cox, 1957, Chapter 14) and John and Quenouille (1977, Chapter 13).

COMPONENT VI: Interpreting and Reporting the Results

Finney (1960, pp.13-14) wrote that

> Too often a statistician imagines that his work is complete when he has designed an experiment and has later computed the analysis of variance (or other appropriate procedure) for the results. It cannot be too strongly emphasized that his responsibility extends further, and that he must play his part, of course in close association with the experimenter, in the interpretation of the statistical analysis. A bare statement of an analysis of variance, with tests of significance and standard errors of means, is no more than the scaffolding around which a report is to be constructed... .

Likewise Yates and Healy (1964, p.205) complained of

> reports on experiments such as "the Z is not significant" without any numerical results.

Despite such clear pronouncements from such eminent authorities, the interpretation of statistical analyses has been badly neglected in teaching of the design and analysis of experiments. As I wrote elsewhere (Preece, 1982a, Section 7):-

> Many academics consider it to be beneath their dignity to provide students with guidance on report-writing; a report written by a student is awarded good marks if it is considered to seem to suggest that the student understands his subject, even if the expression is incoherent and almost devoid of syntax. The subject of the Design of analysis of Experiments has suffered accordingly... .

Skill in writing reports on experiments requires a good feel for clear, concise prose, and ability in both tabular and graphical presentation of means and their standard errors. The writer should know how many significant figures to quote in means, regression coefficients and standard errors, and should be no more shy of using "Plain Numbers" than Civil Servants should be of "Plain Words". As I complained before (Preece, 1982a, Section 7):

> Time and again, when the reader wants to know that, say, Magnaquat made the treated plants grow x ± y cm more than untreated plants, he is fobbed off with "The application of Magnaquat caused a marked increase in plant height" (or "pronounced increase", or "considerable increase", or "notable increase").

It is pleasant to be able to note, in conclusion, that good report writing is encourged by part of "Principle 3" of the Council for National Academic Awards (CNAA, 1979, p.10, paragraph 3.5):-

The programme of studies must enco[illegible] communication and logical argument.

Acknowledgement

This paper was written whilst the author [illegible] Corps) of the Overseas Development Administration (ODA).

References

[1] Aitkin, M. (1981) Response to Query on "Regression models for repeated measurements". *Biometrics*, 37, 831-832.

[2] Anscombe, F.J. (1948) The validity of comparative experiments (with Discussion). *J.Roy.Statist.Soc.*, 111, 181-211.

[3] Chew, V. (1976) Comparing treatment means: a compendium. *Hort Science*, 11, 348-357.

[4] CNAA (1979) *Principles and Regulations for the Award of the Council's First Degrees and Diploma of Higher Education*. London: Council for National Academic Awards.

[5] Cochran, W.G. and Cox, Gertrude M. (1957) *Experimental Designs*, 2nd edition. New York: Wiley. [Available with soft covers].

[6] Cox, D.R. (1958) *Planning of Experiments*. New York: Wiley. [Available with soft covers].

[7] Cox, D.R. (1977) The role of significance tests. *Scand.J.Statist.*, 4, 49-70.

[8] Das, M.N. and Giri, N.C. (1979) *Design and Analysis of Experiments*. New Delhi: Wiley Eastern

[9] Dyke, G.V. (1974) *Comparative Experiments with Field Crops*. London: Butterworths.

[10] Finney, D.J. (1955) *Experimental Design and its Statistical Basis*. Univ. Chicago Press.

[11] Finney, D.J. (1960) *An Introduction to the theory of Experimental Design*. Univ. Chicago Press. [Available with soft covers].

[12] Fisher, R.A. (1925) *Statistical Methods for Research Workers*. Edinburgh: Oliver and Boyd.

[13] Fisher, R.A. (1926) The arrangement of field experiments. *J.Min.Agric.*, 33, 503-513.

[14] Fisher, R.A. (1935) *The Design of Experiments*. Edinburgh: Oliver and Boyd.

[15] Gomez, Kwanchai A. and Gomez, A.A. (1976) *Statistical Procedures for Agricultural Research, with Emphasis on Rice*. Los Banos, Philipines: Internat. Rice Res. Inst.

[16] Goulden, C.H. (1952) *Methods of Statistical Analysis*, 2nd edition. New York: Wiley.

[17] [illegible]

[18] [illegible] (1978) Statistical Checklist 1: Design of Experiments. Cambridge: Inst. Terrestrial Ecology, Natural Environment Res. Council.

[19] John, J.A. and Quenouille, M.H. (1977) Experiments: Design and Analysis, 2nd edition. London: Griffin.

[20] Lewis, D.G. (1968) Experimental Design in Education. Univ. London Press.

[21] Little, T.M. and Hills, F.J. (1978) Agricultural Experimentation: Design and Analysis. new York: Wiley.

[22] Lutz, W. nd Nimmo, I.A. (1977) The inadequacy of statistical significance. Europ. J. Clinical Investigation, 7, 77-78.

[23] Martynov, S.P. and Krupnov, V.A. (1980) Estimation of the significance of differences between varieties. (In Russian) Tsitologiya i Genetika, 14, 83-87. [English translation: Cytology and Genetics, 14, 73-77].

[24] Mead, R. (1982) Contribution to Discussion of "Regression diagnostics, transformations and constructed variables" by A.C. Atkinson. J.Roy.Statist.Soc. B, 44, 30.

[25] O'Neill, R. and Wetherill, G.B. (1971) The present state of multiple comparison methods (with Discussion). J.Roy.Statist.Soc. B, 33, 218-250.

[26] Pearce, S.C. (1982) Review of "Biometry: The Principles and Practice of Statistics in Biological Research, 2nd edition" by R.R. Sokal and J.F. Rohlf, and of "Statistical Tables, 2nd edition" by J.F. Rohlf and R.R. Sokal. Times Higher Ed. supp., No.487 (March 5), 17.

[27] Pocock, S.J. (1982) Statistical aspects of clinical trial design. Statistician, 31, 1-18.

[28] Preece, D.A. (1982a) The design and analysis of experiments: What has gone wrong? Utilitas Mathematica (to appear)

[29] Preece, D.A. (1982b) La biometrie: pas rites mais science. Proc. 11th Internat. Biometric Conf. (to appear)

[30] Raktoe, B.L., Hedayat, A. and Federer, W.T. (1981) Factorial Designs. New York: Wiley.

[31] Rowell, J.G. and Walters, D.E. (1976) Analysing data with repeated observations on each experimental unit. J.Agric.Sci., Cambridge, 87, 423-432.

[32] Taube, A. (1980) Significance, importance and equality - three basic concepts in the analysis of a difference. Upsala J.Med.Sci., 85, 97-102.

[33] Wade, O.L. and Waterhouse, J.A.H. (1977) Significant or important? Brit.J.Clin.Pharmacol., 4, 411-412.

[34] Yates, F. (1937) The design and analysis of factorial experiments. Tech. Commun. No. 35 of the Imperial Bureau of Soil Science (subsequently Commonwealth Bureau of Soils), Harpenden: Imperial Bureau of Soil Science (now Farnham Royal: Commonwealth Agricultural Bureaux).

[35] Yates, F. (1964) Sir Ronald Fisher and the design of experiments. Biometrics, 20, 307-321.

[36] Yates, F. and Healy, M.J.R. (1964) How should we reform the teaching of statistics? J.Roy.Statist.Soc. A, 127, 199-210.

[illegible]

[illegible]

INTRODUCTION

The basic principles of experimental design, as it is usually taught today, both to students specialising in statistics and to those in other disciplines, were established nearly fifty years ago. For many statisticians and users of statistics, the major textbook on Design is still Cochran and Cox (1957), written before the advent of the computer. During the last twenty-five years, our computational habits have changed radically, and it is important to ask whether the fundamental ideas of experimental design remain unchanged, or whether these ideas were influenced by the computational environment in which they were developed.

Do we know the answers to the computer-orientated experimenter who says "There is no need to plan the allocation of treatments to experimental units since we can use the programs for general linear models to sort out the effects of treatments from those of the various environmental factors which we can recognise most effectively after the experiment has begun?" Or the more subtle version, which says "Just try out any treatment allocation on the computer and if the standard errors look OK go ahead and use that allocation."

I shall not be talking about design by computer, but about the teaching of design liberated by the computer from restrictions imposed by the need to analyse the data. I shall attempt to consider not only the teaching of design now, but into the future.

THE GENERAL PRINCIPLES OF DESIGN AND ANALYSIS

Much of what passes for the statistical theory of experimental design is really mathematics and not statistics. The matrix theory of the general linear model in terms of generalised inverses involves no statistical ideas, no implications for the practical collection and interpretation of quantitative data. The extensive theory of partially balanced designs is really only a branch of mathematics. Nevertheless, there are statistical principles which have become submerged in both the mathematics and the details of standard designs, and these principles are what should be taught.

The fundamental principle for the analysis of experimental data is that of fitting sets of parameters, representing the effects of applied treatments, added sequentially to a basic model representing the environmental variation between experimental units. The interpretation of results starts from the analysis of variance (or deviance) which shows the division of the variation between observations into component sums of squares. The most important initial information for interpretation of results is the extra sum of squares due to fitting the latest terms added to the model, combined with the variance-covariance matrix of parameter estimates.

The model and its implications are exemplified most clearly by computer programs, most notably GLIM, which simply allow the addition of sets of terms to the model, calculating the extra sum of squares and providing the variance-covariance matrix. In addition, GLIM recognises the number of distinguishable contrasts it is being required to estimate. The use of GLIM, or some similar program, as a component of an advanced course on Experimental Design is, I believe, essential in the teaching of experimental design. Sometimes GLIM can provide worrying results because of its coding of parameter information, but I have found that the sorting out of apparent paradoxes invariably leads to greater insight into experimental design ideas. One example of the potential of GLIM to worry the innocent user is the analysis of the Latin Square with a missing row shown in Figure 1. If the model

$$Y_{ij} = t_{k(ij)} + r_i + c_j + \epsilon_{ij}$$

is fitted, the terms in the model being written in the order of fitting then the standard errors of the five treatment effect estimates given by GLIM, $\hat{t}_k$ are not all the same. The reason for fitting the treatment effects first was to obtain distinct estimates for all five treatment means, by avoiding constraints on the t_k and gives a clue as to why an obviously balanced design appears unbalanced. A further clue is that the treatment effect for which the standard error is different from the other four is treatment B.

A	D	B	C	E
C	E	A	D	B
D	B	E	A	C
E	A	C	B	D

Figure 1: Latin Square with a missing row

The ideas of statistical design which I believe to be of primary importance are:

(i) Replication
(ii) Blocking
(iii) Randomisation
(iv) Orthogonality
(v) Factorial Treatment Structure
(vi) Confounding
(vii) Covariance
(viii) Quantitative Treatment Factors

Of these, I shall not consider in this paper the ideas of replication, randomisation and factorial structure. This does not imply any lack of devotion to the ideas, each of which figure prominently in any course on experimental design which I teach, but that the presentation of their ideas is not, in my view, essentially altered by the advent of computers.

BLOCKING

The development of the idea of blocking in experimental design depended heavily on the requirement of being able to analyse the results from a blocked experiment. This meant that the Least Squares equations must have a considerable degree of pattern. The sequence of introduction of new design ideas displays a gradual move away from the very simplest form of least squares equations (orthogonality) through balance and on to various patterns of partial balance. The designs thus developed were:

(a) Randomised Block Design
(b) Balanced Incomplete Block Design
(c) Latin Square Designs
(d) Graeco-Latin Square Designs
(e) Partially balanced incomplete block designs
    (i) Lattice Designs
    (ii) Cyclic Designs

It is important to realise that the requirement, per se, of a simple solution for least squares equations is no longer valid. Therefore, the clever tricks for solving sets of equations, represented by the various forms of partial balance, are no longer needed. The relatively simple PBIBD's must therefore be justified as superior to other incomplete block designs in terms of precision of results, not of mathematical nicety.

The importance of orthogonality can now be reassessed. It is that the different sets of effects in the model can be assessed independently without the implicit ambiguity of the non-orthogonal design. The separation of treatment and block effects in the randomised block design is an ideal. However, it is an ideal which in some forms of experiment, for example, clinical trials (Pocock and Simon, 1975) is extremely difficult to achieve. And it is an ideal that is easily disturbed by missing observations. Therefore, while accepting it as an ideal, we must consider how to design experiments when the ideal is unachievable.

In teaching experimental design at all levels, I believe in starting with a problem in which the ideal, orthogonal solution is impossible. The exercise, TOMATO, set to students, and which will be used and discussed in the workshop associated with this invited paper session is described by Mead, Pollock and Ross-Parker (1979). Briefly, a set of twelve glasshouse compartments in two natural groups of six compartments each is available for the experiment. An experimental treatment is necessarily applied to a complete compartment so that there are twelve available experimental units in two blocks of six. The problem presented to the students is to find the best treatment combination from:

2 levels of Heating,
2 levels of Lighting,
2 varieties of tomato.

The problem then is to compare eight treatment combinations in two blocks of six experimental units each.

For students already indoctrinated with the ideas of Randomised Blocks or Confounded Designs, this can be a very worrying problem. Many such students attempt to decide which two treatment combinations to omit so that they can use a Randomised Block Design. However, the logic of choosing a design is very straightforward. I shall leave detailed discussion to the workshop session, but with eight treatments in two blocks of six, it is inevitable that four treatments will be replicated and four not. The replication provides the basis for comparing all eight treatments, and the analysis follows simply in terms of the comparisons within each block.

Having, hopefully, disturbed the presumptions of the students, I then get them to 'solve' in class a sequence of more or less complex design problems. The purpose of this process is to set up the need for some principles for the choice of particular design allocation patterns. The problems can be solved as easily by non-specialist students as by specialist students though only the specialists can follow the mathematics of the theory, the study of which is a natural consequence of the attempts to solve the problems. The ideas are presented in a very intuitive way. The intuition which I seek to develop is an understanding of the statistical theory of experimental design. The mathematics provides further justification if it is needed.

The sequence of problems, all for the comparison of six treatments, all comparisons being assumed equally important, is as follows:

(i) Four blocks of six units each.

Obviously a simple problem to start with, once the students have realised there are no tricks.

(ii) Six blocks of five units each.

Symmetry leads very rapidly to the solution of omitting one treatment in each block, each treatment being omitted from just one block.

(iii) Six blocks of four units each.

It becomes apparent fairly rapidly that there is no satisfying symmetric solution. Two treatments must be omitted from each block. It can readily be seen that there are more than six possible pairs. The students rapidly arrive at a solution of the form shown in Figure 3, although they may get diverted by attempting to write the treatments into the blocks in the order ABCD/EFAB/CDEF/ ABCD/EFAB/CDEF/

(iv) Ten blocks of three units each.

Once they have appreciated that the idea of pairs of treatments occurring together in a block is crucial, and this comes from TOMATO and the solution to problems (ii) and (iii), students come to appreciate that a design with all pairs of treatments occurring once together in a block is possible. And that the construction is almost inevitable.

(v) Six blocks of seven units each.

Again, symmetry rules.

(vi) Three blocks of five units each.

Four blocks of four units each.

Three blocks of three units each.

This, although not solvable in terms of providing equality of precision for all comparisons, is seen to be a genuine problem and to motivate discussion of principles.

For students with the necessary mathematical background, it is now appropriate to consider the least squares equations for a general block-treatment design, and to derive the result due to Pearce (1964) for balance of treatment comparisons. If n_{ij} = the number of observations for treatment j in block i, then a sufficient condition for the equal precision of estimates of treatment differences, for all pairs of treatments, is that the sum of weighted concurrences,

$$\sum_i (n_{ij} n_{ij'} / (\sum_j n_{ij})),$$

should be invariant over all treatment pairs (j, j').

Now although this condition for the balance of all treatment comparisons is clear, it is not clear what criterion should be used in cases when it is not

Block 1	A	B	C	D
Block 2	B	C	D	E
Block 3	C	D	E	F
Block 4	D	E	F	A
Block 5	E	F	A	B
Block 6	F	A	B	C

Figure 2: Six Treatments compared in 6 blocks of 4 units each

possible for the sum of weighted concurrences to be the same for all treatment pairs. Studies of the precision of treatment comparisons in unbalanced designs appear to suggest that the variances of treatment differences are as closely related to the unweighted sum of concurrences, $\sum_i n_{ij} n_{ij'}$, as they are to the weighted sum of concurrences. Further research is needed here.

The inequality of precision in some of the examples should lead to a discussion of why equality of precision should be regarded as an ideal. When the treatment comparisons are markedly unequally precise, then the simplest reliable criterion appears to be the unweighted sum of concurrences. By discussing a range of examples, the effectiveness of this criterion can be demonstrated.

To summarise the ideas I believe should be taught about blocking:

(i) The relevance of orthogonality is changed though still important for avoiding ambiguity.
(ii) Balance and near balance.
(iii) Simple properties of the classical BIB designs.
(iv) The practical desirability of unequal precision.

Particular designs which should be discussed include the Randomised Block Design, Balanced Incomplete Block Design, Latin and possibly Graeco-Latin Squares. Designs which should not be taught include Partially Incomplete Block Designs (including Lattice and Cyclic Designs) and Completely Orthogonal Squares.

To complete this section, I shall use a problem I met recently in an advisory consultation to illustrate the kind of thinking that I believe we should be aiming at in our teaching of blocking. The experimenter wanted to compare 5 diets for cattle, and he had available 9 animals for four successive time periods. The experimental units can be represented as in Figure 3(a). When he arrived for the consultation, the experimenter had convinced himself that he would have to use only 4 treatments in two 4 x 4 Latin Squares, ignoring one animal.

		Animals								
		1	2	3	4	5	6	7	8	9
Periods	I	X	X	X	X	X	X	X	X	X
	II	X	X	X	X	X	X	X	X	X
	III	X	X	X	X	X	X	X	X	X
	IV	X	X	X	X	X	X	X	X	X

(a) The structure of units

A	D	B	C	E	C	E	B	A
C	E	A	D	B	E	D	C	B
D	B	E	A	C	A	C	E	D
E	A	C	B	D	B	A	D	C

(b) The proposed design

Figure 3: Comparison of five diets using 4 period for 9 animals

The first amendment to this was to split the 9 animals into a group of 5 and a group of 4. A 4 x 4 Latin Square could be used for the four more important treatments using the group of 4; and a 5 x 5 Latin Square with a missing row, for all five treatments using the group of 5. This is clearly more efficient than the original design and there is no way of getting more information out of the group of 5. But the group of 4 animals could be used to compare all 5 treatments by using a Latin Square with one row and one column missing. One treatment will now be more precisely estimated than the other four, and there may be problems in choosing which treatment that is to be. However, in comparing five treatments, the symmetry or near symmetry of the design ensures its optimality. The resulting design is shown in Figure 3(b). The supremacy of the intermediate design is clearly arguable, there being advantages for each.

COVARIANCE

The primary importance of covariance in the teaching of design has been through its relationship to blocking. The need in many practical situations to choose between several different blocking criteria leads naturally to consideration of how to allow for variation of factors ignored in the blocking structure. The advantages of using quantitative factors as covariates with qualitative factors for blocking is just one consideration that should be discussed. The problem of the use of covariates is probably particularly important in clinical trials, but does also have importance in other disciplines, such as animal production and psychology. A thorough understanding of the principles and practice of covariance analysis is therefore important in teaching experimental design.

A second, and probably even more important, reason for considering covariance analysis in field crop experiments where the plots are spatially arranged is to introduce the idea of adjustment by covariance for local fertility effects. The method of Papadakis (Bartlett, 1978) has been much discussed recently and must be investigated further before it begins to have much impact on experimental design. However, this is another change in analysis which is encouraged by the easy availability of computers and its general implications for design should not be ignored.

One of the implications of the possible use of Papadakis-style adjustments is that, instead of thinking solely in terms of groups of reasonably homogeneous plots, we should think rather about pairs of adjacent plots. An extreme form would involve considering as 'blocks' all pairs of adjacent plots so that each plot will occur, usually, in four blocks. There are many opportunities here for research, but we should not lose sight of the need to introduce ideas based on Papadakis adjustments into our courses on experimental design in the not-too-distant future.

QUANTITATIVE FACTOR LEVELS

Traditionally, the investigation of the effects of varying the level of a quantitative factor has involved the use of three equally spaced levels. The choice of equal spacing implies the use of a quadratic response function for which the equal spacing gives optimal estimation of the position of any optimum level. The use of just three levels is also optimal unless a response function with more than three parameters is to be fitted, when it is obviously inappropriate.

The computer allows us to use, rather more easily, response functions other than the simple polynomials. We can fit many different response functions and we can also consider designs to discriminate between alternative response functions.

Not only does the computer allow us to consider a wide range of response functions, but it enables us, through approximate results or simulations, to consider optimality of designs for various criteria for any particular response function. As yet, the amount of research on the optimal designs for practically relevant criteria is rather limited, although there is a considerable amount of general theory which indicates clearly the advantages for estimation of using the minimum number of levels. There is a need for much more research on optimal or near-optimal designs for alternative criteria, so that some general patterns relevant to the practical design of experiments can be discerned and taught. Until more information is available, the scope for teaching about the choice of quantitative factors levels in experimental design is limited, but I suggest that this should not provide an excuse for ignoring the subject in our teaching. We can show how computers can already provide the information about optimal design for any particular practical problem and, by teaching the rather sparse information currently available, we can perhaps motivate academic statisticians to find the answers to real problems.

CONFOUNDING

Experimenters believe that confounding is a form of black magic. Textbooks, and sometimes even statisticians, reinforce this view. Most computer programs make the analysis of data from confounded experiments easy, since there is no real difficulty in such analyses; but also it is possible to check, using a computer program such as GLIM, the extent to which effects are confounded in quite complex confounded experiments in which the number of levels per factor varies. I think it is therefore time for a completely new look at how we teach confounding. The motivation for this reconsideration is only partially provided by computers, but the almost total mistrust of confounding by those who should be using confounded experiments provides ample justification.

I believe that confounding is usually taught from quite the wrong viewpoint. Instead of considering primarily the negative question of 'which effects should we confound?', we should ask the more positive question, 'which effects will be unconfounded and therefore available?'.

I suggest that the teaching of confounded experiments should start with the consideration of a complex but practical problem. To illustrate this approach, consider a problem in intercropping experimentation on how to grow a mixture of maize and cowpea crops. The initial experiment consists of three randomised blocks of twelve experimental treatments. The treatments are all combinations of three densities of cowpea (D), two cowpea varieties (C) and two maize varieties (M). The experimenter would ideally also like to include three nitrogen levels (N), in his experiment. Quite apart from the difficulty of having no explicit replication, he would not consider this addition seriously since it would involve the abandonment of blocks, or the division of treatment combinations into blocks, which he knows means confounding. Also, most textbooks do not include a confounded 3 x 3 x 2 x 2 design.

However, consider the problem of adding the nitrogen levels to the initial randomised blocks. To avoid the nitrogen effects being confounded with blocks, we must try to have each of the three nitrogen levels occur four times in each block. We must also consider the allocation of nitrogen levels to the twelve combinations in each block. If we ensure that each nitrogen level occurs with each combination of cowpea and maize varieties, then the interactions NC, NM and NCM will also be unconfounded. The nine combinations of the density and nitrogen factors will necessarily be at least partly confounded with differences between blocks.

We have so far talked only about what we would ideally like. In practice, however, there is no difficulty in allocating the nitrogen levels so as to allow unconfounded estimation of all the effects we hoped for. The resulting design is shown in Figure 4. The construction can be traced by considering the restrictions on the placement of N levels in the order of the observations of block 1 and subsequently of block 2, which in fact requires only a cyclic rotation. The restrictions only start to apply from the fifth combination, except that in the absence of other restrictions, we should try as far as possible to have even replication of the D x N combinations.

The design of Figure 4 is, I believe, a sensible practical experimental design, of a type which should be used far more frequently. This form of design should also be taught at all levels, both specialist and non-specialist. If we consider again the impact of the computer on the analysis and design of experiments, then we can observe two benefits. First, a program such as GLIM is able, if the experimenter is in doubt about the properties of his design, to demonstrate which effects are unconfounded. This should, if confounded designs are properly understood, provide more than a checking facility. The second advantage is related to the ideas discussed in the section on blocking. If we consider the D x N interaction, then we have nine treatment combinations displayed in Figure 5(a), which are allocated to blocks following the design shown in Figure 4 as shown in Figure 5(b). Now, if we had been required to allocate nine treatments to three blocks of twelve plots each, then we should have inexorably come to an allocation with each block containing all nine treatments, plus three duplicated treatments, the three duplicated treatments being different in each block - exactly as in Figure 5.

The computer enables the D x N interactions effects to be estimated. These effects are not orthogonal to blocks, though they are partially balanced (a fact which is quite unnecessary to the appreciation of the advantages of the design). The description of the status of the D x N interaction is not clear from the literature. It is clearly not completely confounded, it is certainly

Block I	Block II	Block III
$d_1\ c_1\ m_1\ n_0$	$d_1\ c_1\ m_1\ n_1$	$d_1\ c_1\ m_1\ n_2$
$d_1\ c_1\ m_2\ n_1$	$d_1\ c_1\ m_2\ n_2$	$d_1\ c_1\ m_2\ n_0$
$d_1\ c_2\ m_1\ n_2$	$d_1\ c_2\ m_1\ n_0$	$d_1\ c_2\ m_1\ n_1$
$d_1\ c_2\ m_2\ n_0$	$d_1\ c_2\ m_2\ n_1$	$d_1\ c_2\ m_2\ n_2$
$d_2\ c_1\ m_1\ n_1$	$d_2\ c_1\ m_1\ n_2$	$d_2\ c_1\ m_1\ n_0$
$d_2\ c_1\ m_2\ n_2$	$d_2\ c_1\ m_2\ n_0$	$d_2\ c_1\ m_2\ n_1$
$d_2\ c_2\ m_1\ n_0$	$d_2\ c_2\ m_1\ n_1$	$d_2\ c_2\ m_1\ n_2$
$d_2\ c_2\ m_2\ n_1$	$d_2\ c_2\ m_2\ n_2$	$d_2\ c_2\ m_2\ n_0$
$d_3\ c_1\ m_1\ n_2$	$d_3\ c_1\ m_1\ n_0$	$d_3\ c_1\ m_1\ n_1$
$d_3\ c_1\ m_2\ n_0$	$d_3\ c_1\ m_2\ n_1$	$d_3\ c_1\ m_2\ n_2$
$d_3\ c_2\ m_1\ n_1$	$d_3\ c_2\ m_1\ n_2$	$d_3\ c_2\ m_1\ n_0$
$d_3\ c_2\ m_2\ n_2$	$d_3\ c_2\ m_2\ n_0$	$d_3\ c_2\ m_2\ n_1$

Figure 4: One replicate of a 3 x 3 x 2 x 2 experiment in 3 randomised blocks

	d_1	d_2	d_3
n_0	1	2	3
n_1	4	5	6
n_2	7	8	9

(a) Coding of the nine treatment combinations from N x D

Block I	Block II	Block III
1	4	7
4	7	1
7	1	4
1	4	7
5	8	2
8	2	5
2	5	8
5	8	2
9	3	6
3	6	9
6	9	3
9	3	6

(b) Allocation of N x D combinations to blocks (from Figure 4)

Figure 5: the non-orthogonal arrangement of the ND Interaction in blocks

not unconfounded, and the term 'partially confounded' is otherwise defined. The natural description would seem to be 'partly confounded'.

CONCLUSION

This has been a very personal view of the teaching of experimental design. Doubtless some of the ideas expressed herein will be felt to be too extreme; possibly some will be regarded as too traditional. My primary concern, however, is to dispel the view of experimental design as a rather staid aspect of statistics, where a range of standard designs exist, all of which can be looked up in the appropriate index in Cochran and Cox, or successors to that worthy volume. This view is only justifiable through a wilful ignorance of the impact of computers. The subject of experimental design is alive and exciting, and there is no reason why it should not be so taught at all levels of teaching.

References

[1] Cochran, W.G. and Cox, G.M. (1957) Experimental Designs, 2nd edition, New York, Wiley.

[2] Mead, R., Pollock, K.H. and Ross-Parker, H.M. (1979) A sequence of games useful in teaching experimental design to agricultural students. American Statistician, 33, 70-76.

[3] Pearce, S.C. (1964) Experimenting with blocks of natural size. Biometrics, 20, 699-706.

[4] Pocock, S.J. and Simon, R. (1975) Sequential treatment assignment with balancing for prognostic factors in the controlled clinical trial. Biometrics, 31, 103-115.

TEACHING CONCEPTS OF DESIGN AND ANALYSIS USING REAL AND SIMULATED EXPERIMENTS

J R Elliott, Wilfrid Laurier University and J C Robinson, University of Waterloo

INTRODUCTION

Courses in experimental design usually provide students with a set of possible designs and corresponding methods of analysis but fail to offer them opportunities for using these new "tools" to solve actual problems. This situation is unfortunate since some of these students may eventually be required to give advice to researchers planning experiments without ever having experienced the joys, frustrations, or compromises involved in conducting an experiment. In fact, with the advent of co-operative, internship, and work-study programs, the student is being requested to give advice even sooner than in the past. How can these students, who have never been confronted by time or cost constraints, choice of appropriate factors, or design of follow-up experiments fully appreciate the problems their clients face?

With the recent development of more powerful and more accessible computer hardware and software, this situation is gradually changing. Several computer-oriented texts are now available containing large sets of "real" data (Afifi and Azen, 1972; Draper and Smith, 1981; Gunst and Mason, 1980). These data are typically used not only for examples and exercises but also to provide a unifying theme for the analytical procedures discussed in the text. However, this approach has not as yet permeated the field of experimental design where the use of small "cooked" data sets is still the norm, even in applied courses. Moreover, this "large data-set" approach is most useful for illustrating the analytical techniques available and tends to ignore the strategies involved in data collection.

The latter, although central to any experimental design or sampling course, are often given a superficial treatment. They are discussed and illustrated, but as noted above few students receive practical experience in their use. W.G. Hunter has suggested, in The American Statistician, two possible remedies; namely, providing the students with computer routines for generating data from simulated experiments or having them obtain "real" data by formulating and performing an actual experiment (Hunter, 1977). Emphasizing the latter approach, he has described a number of simple factorial experiments conceived and conducted by students at the University of Wisconsin. While favouring this method, Hunter noted that, although the total number of possible experiments was large, the range of practical experiments of this type was rather limited.

A more recent article by Pollock et al concentrated on data from simulated experiments and described a sequence of "games" developed at the University of Reading which could be conducted either on the computer or in the classroom (Pollock et al, 1979). These simulations - TOMATO, CHICK and SELECT - were used in teaching design to agriculture students and provided an opportunity for these students to experience firsthand the effects of concepts such as randomization, replication, interaction and blocking on the data-gathering process.

In fact, both approaches have considerable merit as teaching aids in design courses and we too have used both in attempting to make statistical concepts more understandable. Although the mini-research project approach more closely approximates reality, it does have the all-too realistic disadvantage of failing to provide useful results on occasion. Primarily for this reason, we have favoured the use of simulated experiments and have attempted to counter the absence of problem formulation by encompassing a wider variety of design types than could be achieved with real experiments. This paper describes in detail the computer package which we have developed for this purpose and outlines its use as a teaching aid. Following more than ten years of combined experience with this package, we believe that it is a unique, extremely versatile, and powerful tool not only for use in experimental design courses, but in regression, sampling, multivariate, and introductory courses as well.

EXPERIMENT SIMULATION PACKAGE

The Experiment Simulation Package or ESP consists of a series of interactive programs or functions, written in the conversational APL language, and currently available for Honeywell Model 6000, DP58 or level 66 computers with a GCOS3 or GCOS8 operating system, or for Xerox Model Sigma 7 computers with a CP-V version COO (or greater) operating system* (Appendix II). The functions model the carbon monoxide, tar, nicotine and hydrogen cyanide deliveries of cigarettes manufactured by the hypothetical "Ruthless Tobacco Company". The yields of these four combustion products are presumably influenced by six factors; namely, amount of chemical additive, cigarette diameter, filter efficiency, filter overwrap, paper porosity, and cigarette weight (Table 1). (For example, see Baskevitch, 1981; Resnik et al, 1978; Robinson and Forbes, 1975). A different response surface is generated for each of the four combustion products, with carbon monoxide, tar, and nicotine deliveries being defined directly in terms of the six factors, while hydrogen cyanide delivery is a simple quadratic function of both carbon monoxide and tar deliveries (Table 2). The shapes of the curves are different, particularly with respect to the numbers

---

* These programs could be easily transferred to SUPERPET or IBM computers as well, and conversion from APL to FORTRAN should be possible.

TABLE 1: FACTORS OF INTEREST AND ASSOCIATED RANGES

A: CHEMICAL ADDITIVE - a substance added to the tobacco in small amounts which influences deliveries by altering the burning rate.

RANGE: 0 to 0.8mg per cigarette

D: DIAMETER - interacts with weight through density and influences deliveries by diluting the smoke and/or altering combustion.

RANGE: 0.73 to 0.78cm

E: FILTER EFFICIENCY - the material comprising the filter influences deliveries by differential removal of varying amounts of solid particles from the smoke and is available in only four types

TYPES: 15%, 30%, 45%, 60% efficient

O: FILTER OVERWRAP - the cork covering the filter is available in only three porosities and influences deliveries by diluting the smoke

TYPES: Standard, Porous, Very Porous
(Coded: 1, 2, 3 respectively).

P: PAPER TYPE AND POROSITY - the paper covering the tobacco can be treated either with citrate or phosphorus salts and is available in three porosities (the time in seconds required to draw 35ml of air through one square centimeter of paper). It influences deliveries by diluting the smoke since the burning rate increases as the paper becomes more porous.

TYPES: Citrate paper - 10, 20, 30 sec porosity
Phosphorus paper - 10, 20, 30 sec porosity
(Coded 1, 2, 3, 4, 5, 6 respectively)

W: WEIGHT - interacts with diameter through density and influences deliveries by diluting the smoke and/or altering combustion.

RANGE: 0.97 to 1.02gm per cigarette

and types of interactions present. (For example, additive and porosity interact for both tar and carbon monoxide, but not for nicotine). Some of the factors have little or no effect on one response, but have a major influence on others. (For example, increasing filter efficiency has no effect on carbon monoxide delivery, but produces a moderate decrease in nicotine delivery and a massive drop in tar delivery). Consequently, the set of levels for the six factors which minimizes a particular response differs slightly from that which optimises other responses. This feature allows the instructor to focus on one response in class or on the assignments without providing too much information on the optimal solution for the remaining responses. There is, however, enough similarity in the four responses so that predictions can be made, to some extent, for one response based on results from another; that is, the responses are positively correlated through a similar reaction to changes in the levels of various factors such as diameter and weight (Dalhamn, 1975).

In addition to these deterministic components, all four responses are subject to several possible sources of variation or "noise" due to environmental effects. These consist of a pure error component, a random machine run effect, a fixed port effect, and a fixed cigarette lot effect (Table 2). The random error and run components consist of normally distributed variables generated using Marsaglia's polar modification of the Box-Muller method (Box and Muller, 1958; Marsaglia and Bray, 1964), whereas the fixed port and cigarette lot components have U-shaped distributions with the maximum effect being at the extremes. The instructor has control over the magnitude of each component and may easily suppress or implement any combination of these sources of variation. In fact, all variation may be suppressed so that the effect of any combination of levels of the six treatment factors can be determined. This is particularly useful for examining the true effect of supposed "optima" obtained in student projects.

Using the program PARAMETERS, the instructor determines the factor or factors which must be specified by the student, imposes restrictions on the size of the experiment, introduces the error components, determines the responses that will be simulated, and establishes a due data (Figure 1(a)). This program, together with the functions RESOURCES, SIMULATE and DESGEN, and the built-in APL function SETLINK, is permanently stored in a protected workspace currently called ESP..ICOTS. All functions are locked for further protection. Through an unprotected workspace called EXPERIMENT, the student is given access to the locked programs RESOURCES, EXPERIMENT and END. The RESOURCES program informs the student of all restrictions imposed on the experiment by the instructor except for the composition of the error component (Figure 1(b)). The EXPERIMENT program copies the main routine, SIMULATE, from the protected workspace, while the END program erases SIMULATE and saves the responses for the student when the simulation is completed. The heart of ESP is the SIMULATE program which performs the actual simulation of the responses according to the conditions set by the PARAMETERS program. SIMULATE requires as input a design matrix which is usually generated by the student prior to running the simulation (Figure 2(a)). However, if such a matrix is unavailable at the time the SIMULATE program is being executed, then the subroutine DESGEN is retrieved from the protected workspace to aid the student in preparing the proper design matrix. In either case, the SIMULATE program conducts extensive checks of the design matrix for unacceptable, missing, or out-of-range factor levels before proceeding with the simulation. It also denies access to any student attempting a simulation after the due date. A typical terminal output from the SIMULATE program is given in Figure 2(b) corresponding to the design matrix in Figure 2(a). Note that a total cost is printed for the experiment, together with production and testing costs associated with each type of cigarette or treatment combination so that an additional restriction can be imposed on the size of the experiment through budget constraints. Detailed cost information is provided in Appendix 1.

USE OF ESP

ESP has been used as a teaching aid in introductory statistics courses, as well as in the more advanced design and regression courses. In the introductory course, or initially in the more advanced course, the levels of all but one factor are fixed, a random error component is introduced and one of the responses, usually carbon monoxide, is selected for study. Students are required to simulate carbon monoxide deliveries for one or more levels of the factor and simple test statistics are calculated by the students based on their individual responses. Familiarity with the use of ESP, and more generally with the entire APL language, is quickly gained in this fashion. (Some lecture time is devoted to teaching APL which is not a prerequisite for the courses). As the design course progresses, first port and then run effects are introduced so that the student learns to use randomization and blocking to minimize or eliminate such sources of bias. By now, the students have been given assignments, both on the design and on the analysis of completely randomized, randomized block, and Latin square experiments. APL programs[*] for the analysis of data from such experiments are also provided in the EXPERIMENT workspace).

Factorial designs are introduced and the carbon monoxide deliveries are examined when first two and, finally, all six factors are varied simultaneously. The 2^n and 3^n series designs, including the concepts of confounding and fractional

---

[*] Program listings are available from the authors.

TABLE 2: RESPONSE SURFACE MODELS

RESPONSE	MATHEMATICAL MODEL
CARBON MONOXIDE (grams)	$CO = 10 + 4(A - C[P])^2 + (3 - O)^2 + C[P](3 - (O - 2)^2)$ $+ 120(1.77 - 0.2C[P] - W/D^2)^2 + \text{Error}^*$ where $C[P] = 0.1, 0.35, 0.6, 0, 0.25, 0.4$
TAR (milligrams)	$Tar = (1 - E)\{12.5 + 5(0.64 - (A - T[P])^2) + 2T[P] + 0.5(O - 3)^2(1 + D^2/W)$ $+ 20(1.77 - 0.2T[P] - W/D^2)^2\} + \text{Error}^*$ where $T[P] = 0.15, 0.2, 0.5, 0, 0.1, 0.6$
NICOTINE (milligrams)	$Nic = \sqrt{1 - E}\{1.028 + 0.25N[P] + 0.1(O - 3)^2(1 + D^2/W)$ $+ 6(1.77 - 02N[P] - W/D^2)^2\} + \text{Error}^*$ where $N[P] = 0, 0.15, 0.4, 0, 0.15, 0.4$
HYDROGEN CYANIDE (micrograms)	$HCN = \left(\frac{CO + 2Tar}{4}\right)^2$

* The error component is determined by multiplying the average value of the response for the particular experiment by the sum of the pure error, machine run, machine port and cigarette lot effects, if any, appropriate to the cigarette being tested. The pure error and machine run effects are obtained by multiplying randomly generated standard normal values by the corresponding coefficients of variation specified through the PARAMETERS program. The port effect is obtained by multiplying the maximum port error specified through the PARAMETERS program by $\left(\frac{2X_p - N_p - 1}{N_p}\right)^4$ where X_p is the number of the port at which the cigarette is being tested and N_p is the total number of ports available (i.e. the "size" of the smoking machine). This simulates a machine with a single centrally-located suction tube and imposes the maximum variability on ports at each end of the machine (i.e., for $X_p = 1, 2, 3$ or $N_p - 2, N_p - 1, N_p$). The cigarette lot effect is determined in a similar manner except that the 10th power is used so that the range of unaffected or marginally affected lots is substantially larger relative to that for ports.

TABLE 2: RESPONSE SURFACE MODELS - continued...

MINIMUM	FACTOR LEVELS AT MINIMUM					
	A	D	E	O	P	W
10g	0	§	†	VP	P10	§
5mg	0.8	§	0.6	VP	P10	§
0.650mg	†	§	0.6	VP	C10 or P10	§
31.8ug	†	§	0.6	VP	P10	§

NOTES:

§ All responses are related to diameter and weight through density, W/D^2, and are minimized for a density of 1.77. This produces an infinite number of optimal solutions and requires the final solution to be determined on the basis of manufacturing costs.

† Carbon monoxide is unaffected by filter efficiency. Cost considerations would dictate 0.15 or 15% as the appropriate choice. Similarly, nicotine is unaffected by additive and attains its minimum for both 10 second porosity citrate and phosphorus papers. Again, cost considerations would suggest 10 second citrate paper and no additive as the most appropriate choices. Hydrogen cyanide is also independent of additive so that no additive would produce the least expensive optimal cigarette.

FIGURE 1: SAMPLE OUTPUTS FROM PARAMETERS AND RESOURCES PROGRAMS

A: OUTPUT FROM PARAMETERS PROGRAM

```
             PARAMETER RESTRICTIONS AND FACTOR SPECIFICATIONS
             ************************************************

EFFECT OF ADDITIVE,DIAMETER,OVERWRAP,POROSITY AND WEIGHT ON CO DELIVERY

RESTRICTIONS ON SIZE OF EXPERIMENT:     (VARIABLE "SIZERES")
***********************************

                         OBS/RUN     RUNS
MAXIMUM NUMBER:            20         10

SPECIFICATIONS FOR DESIGN MATRIX:     (VARIABLES "FACLEV","FACSTUD")
*********************************

FACTORS WHOSE LEVELS MUST BE SPECIFIED BY STUDENTS:

    ADDITIVE    DIAMETER    OVERWRAP    POROSITY    WEIGHT
     PLUS PORT NUMBERS BETWEEN 1 AND 20 INCLUSIVE

FACTORS WHOSE LEVELS HAVE BEEN PRE"SPECIFIED:

  EFFICIENCY:   0.6

SPECIFICATIONS FOR ERROR AND BIAS TERMS:    (VARIABLE "ERRSPEC")
****************************************

EXPERIMENTAL ERROR:   COEFFICIENT OF VARIATION =  10%

MACHINE RUNS:   COEFFICIENT OF VARIATION =  5%

MACHINE PORTS:   MAXIMUM EFFECT =  1%

CIGARETTE LOTS:   NO EFFECT

RESPONSES TO BE SIMULATED AND PRINTED:     (VARIABLE "RESP")
**************************************

    CO


THE ASSIGNMENT DUE DATE IS *****  AUG. 31, 1982 *****


 ESP          SAVED  20:38 MAY 26,'82
```

B: OUTPUT FROM RESOURCES PROGRAM

```
      RESOURCES

EXPERIMENTAL CONDITIONS:
************************

     THE RESTRICTIONS LISTED BELOW HAVE BEEN IMPOSED ON YOUR EXPERIMENT.
PLEASE REMEMBER THEM WHEN FORMULATING YOUR DESIGN. NOTE ESPECIALLY THAT
THE TOTAL NUMBER OF RUNS MAY NOT EXCEED ***   10   *** AND YOU CAN
HAVE AT MOST ***   20   *** OBSERVATIONS IN EACH RUN SO THE
MAXIMUM NUMBER OF OBSERVATIONS IN YOUR EXPERIMENT IS ***   200   ***.
     YOU MUST SPECIFY IN YOUR DESIGN MATRIX THE NUMBER OF THE PORT AT
WHICH EACH CIGARETTE IS TO BE SMOKED. IN THE PRESENT EXPERIMENT, THE RANGE OF
PORT NUMBERS IS ***   1 TO 20   *** INCLUSIVE AND PORT NUMBERS SHOULD
NOT BE REPEATED WITHIN A RUN UNLESS YOU HAVE BEEN INFORMED OTHERWISE.


     CIGARETTE LOT NUMBERS HAVE BEEN RANDOMLY ASSIGNED BY THE LAB, SO
YOU DO NOT NEED TO SPECIFY THEM.


RESTRICTIONS ON FACTOR LEVELS:
******************************

     YOU MUST SPECIFY LEVELS FOR EACH OF THE FOLLOWING FACTORS
IN YOUR DESIGN MATRIX.

FACTOR:   ADDITIVE    DIAMETER    OVERWRAP    POROSITY    WEIGHT
RANGE:    {0,0.8}     {.73,.78}    {1,3}        {1,6}    {.97,1.02}

     THE FOLLOWING FACTORS HAVE BEEN PRE-SET AT THE LEVELS INDICATED AND
SHOULD NOT BE INCLUDED IN YOUR DESIGN MATRIX.

 EFFICIENCY :   0.6


     RESPONSES WILL CONSIST OF ***      CO      ***
DELIVERIES FOR EACH CIGARETTE.


     YOUR DESIGN AND/OR ANALYSIS MUST BE SUBMITTED BY 5:00 P.M. ON
                ***   AUG. 31, 1982   ***.



YOU MAY NOW CREATE YOUR DESIGN MATRIX, BEARING IN MIND THE OBJECTIVES AND
RESTRICTIONS OUTLINED ABOVE, OR YOU MAY BEGIN YOUR SIMULATION IMMEDIATELY
BY TYPING

                    EXPERIMENT
                    SIMULATE

AND RESPONDING TO THE PROMPTS FROM THE PROGRAM. THE SIMULATION ROUTINE
HAS A BUILT-IN FACILITY FOR HELPING YOU TO CONSTRUCT THE DESIGN MATRIX.
```

FIGURE 2

A: DESIGN MATRIX FOR SIMULATION

(RUN 1: ROWS 1-16 RUN 2: ROWS 17-32)

ADDITIVE	DIAMETER	OVERWRAP	POROSITY	WEIGHT	PORT NO
0.00	0.730	1	2	0.970	15
0.80	0.780	1	2	0.970	8
0.80	0.730	3	2	0.970	17
0.80	0.730	1	5	0.970	16
0.80	0.730	1	2	1.020	6
0.00	0.780	3	2	0.970	11
0.00	0.780	1	5	0.970	13
0.00	0.780	1	2	1.020	5
0.00	0.730	3	5	0.970	14
0.00	0.730	3	2	1.020	3
0.00	0.730	1	5	1.020	12
0.80	0.780	3	5	0.970	7
0.80	0.780	3	2	1.020	9
0.80	0.780	1	5	1.020	18
0.80	0.730	3	5	1.020	10
0.00	0.780	3	5	1.020	4
0.80	0.730	1	2	0.970	17
0.00	0.780	1	2	0.970	5
0.00	0.730	3	2	0.970	14
0.00	0.730	1	5	0.970	8
0.00	0.730	1	2	1.020	15
0.80	0.780	3	2	0.970	4
0.80	0.780	1	5	0.970	9
0.80	0.780	1	2	1.020	12
0.80	0.730	3	5	0.970	7
0.80	0.730	3	2	1.020	13
0.80	0.730	1	5	1.020	11
0.00	0.780	3	5	0.970	18
0.00	0.780	3	2	1.020	10
0.00	0.780	1	5	1.020	6
0.00	0.730	3	5	1.020	3
0.80	0.780	3	5	1.020	16

B: SAMPLE OUTPUT FROM SIMULATE PROGRAM

```
                              RUTHLESS TOBACCO COMPANY
                              ************************

                                                            INVOICE NO.: 3361717
                                                            DATE:    MAY 27, 1982
TO: RESEARCH BRANCH 251839
ATTENTION: J.R.ELLIOTT

                                                                  COSTS
ADDI-   DIA-    EFFI-    OVER-  POR-              PORT   LOT     OF TEST
TIVE    METER   CIENCY   WRAP   OSITY   WEIGHT     NO.   NO.     SAMPLES      CO(MG)

0.00    0.730   0.60      S     C20     0.970      15            $245.93      20.05
0.80    0.780   0.60      S     C20     0.970       8            $270.47      18.50
0.80    0.730   0.60     VP     C20     0.970      17            $266.33      14.95
0.80    0.730   0.60      S     P20     0.970      16            $275.72      16.90
0.80    0.730   0.60      S     C20     1.020       6            $258.33      23.94
0.00    0.780   0.60     VP     C20     0.970      11            $270.07      12.41
0.00    0.780   0.60      S     P20     0.970      13            $280.68      16.12
0.00    0.780   0.60      S     C20     1.020       5            $262.07      16.81
0.00    0.730   0.60     VP     P20     0.970      14            $275.32      10.92
0.00    0.730   0.60     VP     C20     1.020       3            $257.93      17.53
0.00    0.730   0.60      S     P20     1.020      12            $267.32      19.00
0.80    0.780   0.60     VP     P20     0.970       7            $301.08      12.65
0.80    0.780   0.60     VP     C20     1.020       9            $282.47      13.56
0.80    0.780   0.60      S     P20     1.020      18            $293.08      16.76
0.80    0.730   0.60     VP     P20     1.020      10            $287.72      14.35
0.00    0.780   0.60     VP     P20     1.020       4            $292.68      10.79

0.80    0.730   0.60      S     C20     0.970      17            $256.33      20.48
0.00    0.780   0.60      S     C20     0.970       5            $260.07      18.89
0.00    0.730   0.60     VP     C20     0.970      14            $255.93      10.96
0.00    0.730   0.60      S     P20     0.970       8            $265.32      16.74
0.00    0.730   0.60      S     C20     1.020      15            $247.93      18.87
0.80    0.780   0.60     VP     C20     0.970       4            $280.47      14.31
0.80    0.780   0.60      S     P20     0.970       9            $291.08      16.99
0.80    0.780   0.60      S     C20     1.020      12            $272.47      16.20
0.80    0.730   0.60     VP     P20     0.970       7            $285.72      11.34
0.80    0.730   0.60     VP     C20     1.020      13            $268.33      16.81
0.80    0.730   0.60      S     P20     1.020      11            $277.72      21.08
0.00    0.780   0.60     VP     P20     0.970      18            $290.68      10.69
0.00    0.780   0.60     VP     C20     1.020      10            $272.07      10.36
0.00    0.780   0.60      S     P20     1.020       6            $282.68      14.19
0.00    0.730   0.60     VP     P20     1.020       3            $277.32      14.58
0.80    0.780   0.60     VP     P20     1.020      16            $303.08      11.65
                                                                 -------
                        TOTAL COST OF TEST SAMPLES:        $   8774.40
                        LABORATORY CHARGE FOR TESTING: $   1440.00
                                                           ----------

                        TOTAL COST OF EXPERIMENT:          $ 10214.40
                            ( 32 DETERMINATIONS)
```

C: SAMPLE OUTPUT FROM YATES' ANALYSIS PROGRAM

```
CALCULATIONS FROM YATES ALGORITHM
*********************************

EFFECT OF ADDITIVE,DIAMETER,OVERWRAP,POROSITY AND WEIGHT ON CO DELIVERY.


FACTORS
*******

A ¯ ADDITIVE
D ¯ DIAMETER
O ¯ OVERWRAP
P ¯ POROSITY
W ¯ WEIGHT

          EFFECT                  MEAN SQUARE    SOURCE
          ******                  ***********    ******


         1.34652389               14.50501264    A
        ¯2.35146344               44.23504256    D
        ¯0.05287523                0.02236632    AD
        ¯5.22897135              218.73713124    O
         0.07553895                0.04564906    AO
         0.47182882                1.78097949    DO
         0.60979595                2.97480879    ADO
        ¯1.86477804               27.81917706    P
        ¯0.26173253                0.54803133    AP
         0.46694587                1.74430755    DP
         0.53071160                2.25323840    ADP
         0.12902434                0.13317825    OP
        ¯0.41018827                1.34603531    AOP
         0.05773496                0.02666661    DOP
        ¯0.43157086                1.49002726    ADOP
         0.84766564                5.74829625    W
         0.17952511                0.25783411    AW
        ¯2.12865819               36.24948566    DW
         0.03128599                0.00783050    ADW
         0.57676670                2.66127864    OW
        ¯0.82418173                5.43420425    AOW
        ¯0.22028360                0.38819892    DOW
         0.66440226                3.53144294    ADOW
         0.40679330                1.32384630    PW
         0.05336085                0.02277904    APW
         0.10865603                0.09444906    DPW
        ¯0.11315684                0.10243576    ADPW
        ¯0.39036221                1.21906125    OPW
         0.15264198                0.18639660    AOPW
         0.34829098                0.97045283    DOPW
        ¯0.69515916                3.86597001    ADOPW


TABLE OF FACTOR MEAN RESPONSES:
*******************************

FACTOR                LOW LEVEL          HIGH LEVEL          DIFFERENCE

ADDITIVE             14.93242535         16.27894924          1.34652389
DIAMETER             16.78141901         14.42995557         ¯2.35146344
OVERWRAP             18.22017297         12.99120162         ¯5.22897135
POROSITY             16.53807631         14.67329827         ¯1.86477804
WEIGHT               15.18185448         16.02952011          0.84766564
```

replication, are discussed in detail and illustrated with the simulator. The design matrix, simulated results, and partial analysis for a typical assignment are given in Figure 2. (A program* for performing an analysis of variance using Yates' method is also provided in the EXPERIMENT workspace).

Following consideration of the 2^n series, the students are assigned a "real" data project of the type described by W.G. Hunter (Hunter, 1977). The general factorial design is discussed and illustrated, as are miscellaneous topics such as composite designs, split plot experiments, incomplete block designs, and simple response surface methodology. At this point, the students have enough background to attack a problem based upon the full complexities of the simulator and a major project is assigned requiring each student or team of students to determine the levels of the six factors which will minimise one of the remaining responses; that is, tar, nicotine, or hydrogen cyanide. In larger classes, different responses can be assigned to different teams. However, it is our impression that providing everyone with the same response stimulates competition and improves the quality of the final products. At the end of the projects, a written report of all findings is submitted, sometimes accompanied by a verbal "in-class" presentation. Each separate experiment is described as to choice of design (including rationale), results from analysis of responses, and conclusions regarding the direction of future experiments. Recommendations regarding the optimal cigarette are presented, along with the total cost of experimentation. Project evaluation is based upon use and understanding of design, coherence of the sequence of experiments, proximity of recommended cigarette to the optimum, and economy of experimentation (with respect to total cost and to number of experiments). Typically, the two projects (the real and simulated experiments) account for a substantial proportion of the student's final grade in the course.

DISCUSSION AND CONCLUSIONS

Like most instructors who teach experimental design as a service course, we have attempted to provide our students with a variety of methods for scientific (that is, statistically valid) data collection and analysis. They are presented with a variety of distinct data collection strategies (that is, designs), each of which applies to a slightly different experimental situation and each of which is accompanied by a related but distinct method of analysis. As with D.R. Cox's approach in "Planning of Experiments" (Cox, 1958), our goal is to instill in our students an appreciation of which design is most appropriate for a given experimental setting, stressing at all times that there is really no "uniquely correct" answer but rather a series of tradeoffs between cost and precision (as defined in the broad sense).

We have found it difficult to accomplish this objective using lectures based solely on traditional experimental design texts (Cochran and Cox, 1957; Mendenhall, 1968; Snedecor and Cochran, 1967). Simply rehashing other people's data does not expose the students to the very real problems of factor and design selection that will be encountered when conducting their own experiments. Thus, like Hunter, we have supplemented this text material with a combination of real and simulated experiments. The former allow the students to experience first-hand the joys and trials of data collection and analysis; they must formulate a problem, select relevant responses and suspected influential factors, conduct an experiment, analyze the results, and prepare a coherent written report of their findings. This approach is extremely exciting to the students who become deeply involved in their projects. Yet, by their very nature, such projects are limited in scope, and must not be too large or too complicated. Moreover, they are subject to all the frustrations of the real research projects which they model; in particular, they have an annoying habit of revealing absolutely no statistically significant differences between the levels of the factors selected. Still as noted by Hunter, discussion of such outcomes can be a learning experience as well, provided neither the instructor nor the students place too much emphasis on the actual results of the experiments.

In contrast, with simulated experiments, the student sacrifices the opportunity to formulate a problem and select relevant responses. This is more than balanced, in our opinion, by the versatility through exposure to a wide variety of situations that such simulators in general, and ESP in particular, possess. In addition, the occurrence of missing observations is avoided and the time between selection of a design and production of results is greatly shortened. For these reasons, we have made ESP an integral part of our approach to experimental design. Besides using it for assignments and examples, we have written a text to accompany the course which makes extensive reference to the situation being modelled in illustrating all of the designs considered. This has a unifying influence on the course similar to that achieved with large data sets in multiple regression texts; namely, it provides a common context in which different designs can be compared on a variety of dimensions (for example, efficiency, degree of complexity, or cost). Moreover, the extreme versatility of ESP permits rapid production of data from numerous design situations - simple or complex, single or multi-factor, complete or incomplete block, whole or split plot - by slight adjustment to the model using the PARAMETERS program. This virtually limitless capability for producing material for assignments and examples frees the instructor from an otherwise time-consuming task.

To date, student reaction to the package has been extremely positive. On assignments, many students initially discount the possibility of port or run biases and use completely randomised designs. However, when their classmates achieve a smaller experimental error in fewer observations with randomised block or Latin square designs, most begin to appreciate the sacrifice in precision that accompanies underdesigned experiments. In contrast, the more cautious students complicate their experiments unnecessarily by over-blocking or by using an excessive number of runs and/or factor levels. Frequently they explore one factor at a time and thus miss important interactions; failing to obtain significant effects or finding that the results of an experiment contradict previous conclusions, they seek advice from their instructor. Discussion of these "failed" experiments individually or in class is beneficial for illustrating, in a concrete manner, the advantage of applying factorial designs with minimal replication early in the experimental program. Moreover, concepts such as confounding and fractional replication, which at first seem somewhat elusive to the student, become valuable tools during the project. Slowly, sometimes painfully, most of the students come to realize that experimental design is not simply a theory but rather a very practical and useful concept.

Students find the simulation package simple to use and appear to derive much personal satisfaction from conducting the project. This positive attitude towards the simulation program seems to stem in part from the ease with which students can relate to the situation being modelled, a feature shared with those described by Pollock et al. Almost everyone, regardless of smoking status or academic discipline, has some knowledge of cigarette construction and the products of tobacco combustion. Terms such as "filter cigarette", "additive", "diameter", "weight", "tobacco", "tar", "nicotine" and "carbon monoxide" are familiar, so that both the factors and the responses have an intuitive appeal. Thus, students can readily appreciate the nature of the factors and reasons why they might influence tobacco combustion. A knowledge of cigarette manufacturing technology is not necessary for students to predict, for example, that increasing filter efficiency should decrease tar and nicotine delivery. Similarly, knowledge of human physiology is not essential for understanding that reducing carbon monoxide, tar, nicotine, or hydrogen cyanide output should produce a less hazardous cigarette. In these many ways, the simulation package serves as a driving force to make the theory of design a reality and obscure concepts more concrete.

References

[1] Afifi, A.A. and Azen, S.P. (1972) Statistical Analysis: A Computer Oriented Approach. New York: Academic Press.

[2] Baskevitch, N. (1981) Continuing Upward Trend in Low Tar and Nicotine Cigarettes. Tabak Journal International, 4, 304-306, 308, 310.

[3] Box, G.E.P. and Muller, M.E. (1958) A note on the generation of random normal deviates. *Annals of Mathematical Statistics*, 29, 610-611.

[4] Cochran, W.G. and Cox, G.M. (1957) *Experimental Designs*, 2nd edition. New York: Wiley.

[5] Cox, D.R. (1958) *Planning of Experiments*. New York: Wiley.

[6] Dalhamn, T. (1975) Chronic Pulmonary Disease - The Etiological Factors in Smoke. *Proceedings of the Third World Conference on Smoking and Health*, 415-426.

[7] Draper, N.R. and Smith, H. (1981) *Applied Regression Analysis*, 2nd edition. New York: Wiley.

[8] Gunst, R.F. and Mason, R.L. (1980) *Regression Analysis and its Application: A Data-Oriented Approach*. New York: Marcel Dekker.

[9] Hunter, W.G. (1977) Some Ideas About Teaching Design of Experiments, with 2^5 Examples of Experiments Conducted by Students. *The American Statistician*, 31, 12-17.

[10] Marsaglia, G. and Bray, T.A. (1964) A Convenient Method for Generating Normal Variables. *Society for Industrial and Applied Mathematics: Review*, 6, 260-264.

[11] Mendenhall, W. (1968) *Introduction to Linear Models and the Design and Analysis of Experiments*. Belmont: Duxbury Press.

[12] Pollock, K.H., Ross-Parker, H.M. and Mead, R. (1979) A Sequence of Games Useful in Teaching Experimental Design to Agriculture Students. *The American Statistician*, 33, 70-76.

[13] Resnik, F.E., Houck, W.G. *et al* (1978) Factors Affecting Static Burning Rate. *Tobacco Science*, 21, 103-107.

[14] Robinson, J.C. and Forbes, W.F. (1975) The Role of Carbon Monoxide in Cigarette Smoking: 1. Carbon Monoxide Yield from Cigarettes. *Archives of Environmental Health*, 31, 425-434.

[15] Snedecor, G.W. and Cochran, W.G. (1967) *Statistical Methods*, 6th edition. Ames: Iowa State University Press.

APPENDIX I: PRODUCTION AND TESTING COSTS

PRODUCTION COSTS FOR TEST SAMPLES

The standard cigarette contains no additive, is 0.755cm in diameter with a 0.215cm overlap, weighs approximately 0.97gm and is covered with 30 second porosity citrate paper. The filter has a 15% efficiency and a standard overwrap.

Cigarettes of each type are manufactured in lots of 10,000 and the standard cigarette costs $233.00 per lot to produce (including government tax).

Additive costs $0.50 per gram and must be applied by hand at a flat rate of $10 per lot.

Changing the diameter from the standard value of 0.755cm increases or decreases cost proportional to the amount and type of paper required.

Filter material of 15% efficiency is included in the basic cost. Increasing the efficiency to 30% adds $5 per lot, to 45% adds $10 per lot, and to 60% adds $20 per lot.

Using porous or very porous overwraps adds $5 or $10, respectively, to the base cost per lot.

Citrate paper, regardless of porosity, adds nothing to the base cost, whereas phosphorus paper increases the cost by $30 per lot for 10 sec, $20 per lot for 20 sec, and $10 per lot for 30 sec porosity.

Weights above 0.97gm per cigarette increase the cost by $4 per kg of additional tobacco required.

LABORATORY COST FOR TESTING CIGARETTES

The first run costs $500 and yields tar deliveries whether they are requested or not. Carbon monoxide and nicotine determinations each add $300 to the base cost, while hydrogen cyanide determination adds a further $500. Thus, for example, a single run for carbon monoxide delivery only would cost $800 and a single run in which all four responses were monitored would cost $1600.

Each additional run adds 80% of the cost of the first run to the total.

APPENDIX II: PROGRAM LISTINGS FOR EXPERIMENT SIMULATION PACKAGE

A. PROGRAM PARAMETERS

```
    ∇ PARAMETERS;LABEL;LOC;NP;I
[1]    '
RESTRICTIONS ON SIZE OF EXPERIMENT:
***********************************

MAXIMUM NUMBER OF:
     OBSERVATIONS PER RUN?'
[2]    SIZERES←⎕
[3]    '       RUNS?'
[4]    SIZERES←SIZERES,⎕
[5]    '
SPECIFICATION OF FACTOR LEVELS:
********************************

(TYPE S IF FACTOR TO BE SPECIFIED BY STUDENT; OTHERWISE, SPECIFY LEVEL)

ADDITIVE?     (0.0 TO 0.8 MG PER CIGARETTE)'
[6]    FACLEV←⎕
[7]    'DIAMETER?     (0.73 TO 0.78 CM)'
[8]    FACLEV←FACLEV,⎕
[9]    'EFFICIENCY?       (.15,.30,.45,.60)'
[10]   FACLEV←FACLEV,⎕
[11]   'OVERWRAP?      (STANDARD,POROUS,VERY POROUS''''CODED: 1,2,3 )'
[12]   FACLEV←FACLEV,⎕
[13]   'POROSITY?      (C10,C20,C30,P10,P20,P30''''CODED: 1,2,3,4,5,6 )'
[14]   FACLEV←FACLEV,⎕
[15]   'WEIGHT?      (0.07 TO 1.02 GM PER CIGARETTE)'
[16]   FACSTUD←S=FACLEV←FACLEV,⎕
[17]   '
SPECIFICATION OF EXPERIMENTAL ERROR AND BIAS TERMS:
***************************************************

(TYPE 0 TO SUPPRESS EXPERIMENTAL ERROR OR SPECIFIC BIAS TERMS)

EXPERIMENTAL ERROR:      (NORMAL RANDOM VARIABLE)
     COEFFICIENT OF VARIATION (%)?   (USUALLY 5¨10 %)'
[18]   ERRSPEC←⎕
[19]   'MACHINE RUNS:      (NORMAL RANDOM VARIABLE)
     COEFFICIENT OF VARIATION (%)?   (USUALLY 5¨10 %)'
[20]   ERRSPEC←ERRSPEC,⎕
[21]   'MACHINE PORTS:      (U¨SHAPED FIXED EFFECT)
     NUMBER OF PORTS PER RUN?      (USUALLY 5¨20)'
[22]   →(0=¯1↑FACSTUD←FACSTUD,(0≠¯1↑FACLEV←FACLEV,⎕))/L0
[23]   '      MAXIMUM EFFECT (%)?   (PORT EFFECT IS PROPORTIONAL
                              TO 4TH POWER OF DISTANCE FROM CENTER)'
[24]   ERRSPEC←ERRSPEC,⎕
[25]   →L1
[26]   L0:ERRSPEC←ERRSPEC,0
[27]   L1:'CIGARETTE LOTS:      (U¨SHAPED FIXED EFFECT)
     NUMBER OF CIGARETTE LOTS?   (USUALLY 100¨1000)'
[28]   →(0=¯1↑FACSTUD←FACSTUD,(0≠¯1↑FACLEV←FACLEV,⎕))/L3
[29]   '      MAXIMUM EFFECT (%)?   (LOT EFFECT IS PROPORTIONAL TO 10TH
                                        POWER OF DISTANCE FROM CENTER)'
[30]   ERRSPEC←ERRSPEC,⎕
[31]   →L4
[32]   L3:ERRSPEC←ERRSPEC,0
[33]   L4:'
RESPONSES TO BE SIMULATED:
**************************

(TYPE S IF RESPONSE IS TO BE SIMULATED; OTHERWISE, TYPE N)

CARBON MONOXIDE?  (10¨20 GM)'
[34]   RESP←1-'N'∊⍞
[35]   'TAR?  (5¨20 MG)      '
[36]   RESP←RESP,1-'N'∊⍞
[37]   'NICOTINE?'
[38]   RESP←RESP,1-'N'∊⍞
[39]   'HCN?'
[40]   RESP←RESP,1-'N'∊⍞
[41]   '
ASSIGNMENT DUE DATE:       MONTH/DAY/YEAR      (E.G. 10 25 83)
*********************'
[42]   DUEDATE←⎕
[43]   '
ENTER A TITLE FOR THE PARAMETER SETTINGS, THEN ADVANCE PAPER TO A FRESH
PAGE AND PRESS THE RETURN BUTTON'
[44]   LABEL←⍞
[45]   '

              PARAMETER RESTRICTIONS AND FACTOR SPECIFICATIONS
              ************************************************

',LABEL
[46]   TEMP←1↑0
```

BEST COPY AVAILABLE

A: PARAMETERS (CONT'D)

```
[47]  '
RESTRICTIONS ON SIZE OF EXPERIMENT:      (VARIABLE "SIZERES")
***********************************

                          OBS/RUN      RUNS
MAXIMUM NUMBER:';,'I10,I11'TEMP ⍕ 2ρSIZERES
[48]  ⍎')ERASE TEMP'
[49]  '
SPECIFICATIONS FOR DESIGN MATRIX:      (VARIABLES "FACLEV","FACSTUD")
**********************************

FACTORS WHOSE LEVELS MUST BE SPECIFIED BY STUDENTS:
     '
[50]  ,(LOC←FACSTUD[⍳6])⌿(⍳1)LABEL←6 13ρ'    ADDITIVE      DIAMETER  EFFICIENCY    OVERWRAP      POROSITY      WEIGHT  '
[51]  →(~FACSTUD[7])/L5
[52]  '       PLUS PORT NUMBERS BETWEEN 1 AND ';FACLEV[7];' INCLUSIVE'
[53]  →(~FACSTUD[8])/L5
[54]  '       AND CIGARETTE LOT NUMBERS BETWEEN 1 AND ';FACLEV[8];' INCLUSIVE'
[55]  L5:→(6=+/LOC)/L6
[56]  '
FACTORS WHOSE LEVELS HAVE BEEN PRE-SPECIFIED:
     '
[57]  I←1
[58]  NP←ρLOC←(1-LOC)/⍳6
[59]  L7:→(L8,L8,L8,LO,LP,L8)[LOC[I]]
[60]  LO:'      OVERWRAP:    ';(3 11ρ'STANDARD   POROUS     VERY POROUS')[FACLEV[4];]
[61]  →1+L8
[62]  LP:'      POROSITY:    ';(6 3ρ'C10C20C30P10P20P30')[FACLEV[5];]
[63]  →L8+1
[64]  L8:LABEL[LOC[I];];':    ';FACLEV[LOC[I]]
[65]  →(NP≥I←I+1)/L7
[66]  L6:'
SPECIFICATIONS FOR ERROR AND BIAS TERMS:    (VARIABLE "ERRSPEC")
***************************************

EXPERIMENTAL ERROR:   COEFFICIENT OF VARIATION =  ';ERRSPEC[1];' %'
[67]  →(0=ERRSPEC[2])/L61
[68]  '
MACHINE RUNS:   COEFFICIENT OF VARIATION =  ';ERRSPEC[2];' %'
[69]  →L611
[70]  L61:'
MACHINE RUNS:   NO EFFECT'
[71]  L611:→(0=ERRSPEC[3])/L62
[72]  '
MACHINE PORTS:   MAXIMUM EFFECT =  ';ERRSPEC[3];' %'
[73]  →L621
[74]  L62:'
MACHINE PORTS:   NO EFFECT'
[75]  L621:→(0=ERRSPEC[4])/L63
[76]  '
CIGARETTE LOTS:   MAXIMUM EFFECT =  ';ERRSPEC[4];' %'
[77]  →L0
[78]  L63:'
CIGARETTE LOTS:   NO EFFECT'
[79]  L0:'
RESPONSES TO BE SIMULATED AND PRINTED:     (VARIABLE "RESP")
**************************************

     ';,RESP/(3 4ρ'CO      TAR     NICOTINE    HCN'
[80]  LABEL←12 5ρ' JAN. FEB. MAR. APR.  MAY JUNE JULY AUG.SEPT. OCT. NOV. DEC.'
[81]  '

THE ASSIGNMENT DUE DATE IS ***** ';LABEL[DUEDATE[1];];' ';DUEDATE[2];', 19';DUEDATE[3];' *****'
[82]  ⍎')GROUP GLOBVARS'
[83]  ⍎')GROUP GLOBVARS SIZERES FACLEV FACSTUD ERRSPEC RESP DUEDATE'
[84]  '


     '
[85]  ⍎')SAVE'

     ∇
```

BEST COPY AVAILABLE

D: PROGRAM RESOURCES

```
    ∇ RESOURCES;LABEL;[illegible]
[1]   [illegible] ALLOWANCE'
[2]   '

EXPERIMENTAL CONDITIONS:
************************

     THE RESTRICTIONS LISTED BELOW HAVE BEEN IMPOSED ON YOUR EXPERIMENT.
PLEASE REMEMBER THEM WHEN FORMULATING YOUR DESIGN. NOTE ESPECIALLY THAT
THE TOTAL NUMBER OF RUNS MAY NOT EXCEED ***   ';SIZERES[1];'   *** AND YOU CAN
HAVE AT MOST ***   ';SIZERES[2];'   *** OBSERVATIONS IN EACH RUN SO THE
MAXIMUM NUMBER OF OBSERVATIONS IN YOUR EXPERIMENT IS ***   ';[illegible];'   ***.'
[3]   LABEL←[illegible]'   ADDITIVE     DIAMETER   EFFICIENCY    OVERWRAP    POROSITY    WEIGHT       [illegible]     [illegible]
      [illegible]        [illegible]         [illegible]    [illegible]'
[4]   →(~FACSTAR[7])/L1
[5]   '      YOU MUST SPECIFY IN YOUR DESIGN MATRIX THE NUMBER OF THE PORT AT
WHICH EACH CIGARETTE IS TO BE SMOKED. IN THE PRESENT EXPERIMENT, THE RANGE OF
PORT NUMBERS IS ***   1 TO ';FACLEVEL[7];'   *** INCLUSIVE AND PORT NUMBERS SHOULD
NOT BE REPEATED WITHIN A RUN UNLESS YOU HAVE BEEN INFORMED OTHERWISE.
     '
[6]   →(~FACSTAR[8])/L2
[7]   '
     YOU MUST SPECIFY IN YOUR DESIGN MATRIX THE NUMBER OF THE LOT FROM
WHICH THE CIGARETTE WAS CHOSEN. IN THE PRESENT EXPERIMENT, THE RANGE OF
LOT NUMBERS IS ***   1 TO ';FACLEVEL[8];'   *** INCLUSIVE. LOT NUMBERS MAY BE
REPEATED IF NECESSARY, EVEN WITHIN THE SAME TYPE OF CIGARETTE, BUT THIS
IS NOT ADVISEABLE.'
[8]   →L3
[9]   L2:'
     CIGARETTE LOT NUMBERS HAVE BEEN RANDOMLY ASSIGNED BY THE LAB, SO
YOU DO NOT NEED TO SPECIFY THEM.
     '
[10]  →L3
[11]  L1:'
     BOTH PORT AND CIGARETTE LOT NUMBERS HAVE BEEN RANDOMLY ASSIGNED
BY THE LAB, SO YOU DO NOT NEED TO SPECIFY THEM.
     '
[12]  L3:'

RESTRICTIONS ON FACTOR LEVELS:
******************************* '
[13]  '
     YOU MUST SPECIFY LEVELS FOR EACH OF THE FOLLOWING FACTORS
IN YOUR DESIGN MATRIX.
     '
[14]  'FACTOR: ';[illegible]
[15]  'RANGE: ';[illegible]
[16]  [illegible]
[17]  [illegible]
[18]  '
     THE FOLLOWING FACTORS HAVE BEEN PRESET AT THE LEVELS INDICATED AND
SHOULD NOT BE INCLUDED IN YOUR DESIGN MATRIX.
     '
[19]  I←1
[20]  [illegible]
[21]  [illegible]'     OVERWRAP :     ';[illegible]'STANDARD    POROUS      VERY POROUS'[illegible]
[22]  [illegible]
[23]  [illegible]'     POROSITY :     ';[illegible]
[24]  [illegible]
[25]  [illegible]LABEL[illegible];':    ';[illegible]
[26]  [illegible]
[27]  [illegible]'

     RESPONSES WILL CONSIST OF ***   ';[illegible]   ***
DELIVERIES FOR EACH CIGARETTE.'
[28]  LABEL←[illegible]'JAN. FEB. MAR. APR.  MAY JUNE JULY AUG. SEPT. OCT. NOV. DEC.'
[29]  '

     YOUR DESIGN AND/OR ANALYSIS MUST BE SUBMITTED BY 5:00 P.M. ON
               ***   ';LABEL[DUEDATE[1];];' ';DUEDATE[2];', 19';DUEDATE[3];'   ***.
  '
[30]  '


YOU MAY NOW CREATE YOUR DESIGN MATRIX, BEARING IN MIND THE OBJECTIVES AND
RESTRICTIONS OUTLINED ABOVE, OR YOU MAY BEGIN YOUR SIMULATION IMMEDIATELY
BY TYPING

                  EXPERIMENT
                  SIMULATE

AND RESPONDING TO THE PROMPTS FROM THE PROGRAM. THE SIMULATI
HAS A BUILT IN FACILITY FOR HELPING YOU TO CONSTRUCT THE DESI
  '
[31]  [illegible] ALLOWANCE'
    ∇
```

BEST COPY AVAILABLE

A. PROGRAM EXPERIMENT

```
    ∇ EXPERIMENT
[1]   [illegible]
    ∇
```

B. PROGRAM SIMULATE

```
    ∇ SIMULATE [illegible]
[1]   [illegible]
[2]   [illegible] JAN. FEB. MAR. APR. MAY JUNE JULY AUG. SEPT. OCT. NOV. DEC.[illegible]
[3]   ⍝
[4]   ⍝   CURRENT DATE COMPARED WITH DUE DATE AND PROGRAM
[5]   ⍝   TERMINATED IF TOO LATE.
[6]   ⍝
[7]   [illegible]
[8]   [illegible]
[9]   [illegible]
[10]  [illegible]
[11]  TURF:'MANAGEMENT MADE THE DECISION ON [illegible]
YOUR ADVICE IS A LITTLE LATE.  CLEAN OUT YOUR DESK[illegible]YOU ARE FIRED ***'
[12]  →LERASE
[13]  ⍝
[14]  ⍝   OBTAINING THE DESIGN MATRIX
[15]  ⍝
[16]  [illegible]'IS A DESIGN MATRIX AVAILABLE FOR INPUT?  (YES OR NO)'
[17]  [illegible]
[18]  'INPUT THE DESIGN MATRIX.'
[19]  DESMAT←⎕
[20]  →[illegible]
[21]  ⍝
[22]  ⍝   DESIGN MATRIX "DESMAT" IS NOW CREATED
[23]  ⍝   BY THE SUBPROGRAM "DESIGN" IF NOT AVAILABLE
[24]  ⍝   EXTERNALLY TO "SIMULATE"
[25]  ⍝
[26]  [illegible] DESIGN'
[27]  DESIGN
[28]  ⍎')ERASE DESIGN'
[29]  ⍝
[30]  ⍝   THE DESIGN MATRIX "DESMAT" IS NOW TESTED TO ENSURE
[31]  ⍝   THAT IT CONFORMS TO ALL RESTRICTIONS IMPOSED
[32]  ⍝   BY THE PROGRAM "[illegible]".
[33]  ⍝
[34]  [illegible]
[35]  'THE MAXIMUM NUMBER OF OBSERVATIONS ALLOWED IS ***  [illegible]  ***
***** BE[illegible]CHECK THE DESIGN MATRIX AND TRY AGAIN *****'
[36]  →LERASE
[37]  P[illegible]:'NUMBER OF RUNS?'
[38]  [illegible]
[39]  'YOU ARE ONLY ALLOWED ***  [illegible]  *** RUNS.
***** BE[illegible]CHECK THE DESIGN MATRIX AND TRY AGAIN *****'
[40]  →LERASE
[41]  [illegible]
[42]  'THE TOTAL NUMBER OF OBSERVATIONS IS NOT AN
INTEGRAL MULTIPLE OF THE NUMBER OF RUNS.
***** BE[illegible]CHECK THE DESIGN MATRIX AND TRY AGAIN *****'
[43]  →LERASE
[44]  [illegible]
[45]  'YOU ARE ONLY ALLOWED ***  [illegible]  *** OBSERVATIONS PER RUN.
***** BE[illegible]CHECK THE DESIGN MATRIX AND TRY AGAIN *****'
[46]  →LERASE
[47]  ⍝
[48]  ⍝   THE DESIGN MATRIX  IS NOW IMBEDDED IN AN
[49]  ⍝   MODEL BY A MATRIX CALLED "DMAT" AND IS TESTED
[50]  ⍝   TO VERIFY THAT ALL VALUES ARE IN THE PROPER RANGES.
[51]  ⍝
[52]  [illegible]
[53]  [illegible]
[54]  [illegible] ADDITIVE  DIAMETER EFFICIENCY [illegible]  POROSITY  WEIGHT  [illegible] NO.  LOT NO. '
[55]  [illegible]
[56]  [illegible]
[57]  [illegible]
[58]  [illegible]
[59]  '
AT LEAST ONE OF THE SPECIFIED ***  [illegible]  *** LEVELS IS
OUTSIDE THE ACCEPTABLE RANGE ***  [illegible]  ***.
***** BE[illegible]CHECK THE DESIGN MATRIX AND TRY AGAIN *****'
[60]  →LERASE
[61]  [illegible]
[62]  [illegible]
[63]  [illegible]
[64]  '
AT LEAST ONE OF THE SPECIFIED ***  EFFICIENCY  *** LEVELS IS
NOT 0.[illegible] OR 0.[illegible] , THE ONLY VALUES PERMITTED.
***** BE[illegible]CHECK THE DESIGN MATRIX AND TRY AGAIN *****'
[65]  →LERASE
```

BEST COPY

6. SIMULATE (cont'd)

```
[illegible]
```

BEST COPY

D: SIMULATE (CONT'D)

```
[150] P7:→(~RESP[4])/P8
[151] DMAT←DMAT,[2]HCN←(0.25×CO+2×TAR)÷2
[152] DMAT←DMAT[;⍳9],((0 9)↓DMAT)[;RESP/(1,2,3,3+RESP[3])]
[153] P8:→(4=+/RESP)/P9
[154] ⍎')ERASE  ',,(~RESP)/[1]4 9⍴'        CO       TAR NICOTINE     HCN  '
[155] ⍝
[156] ⍝   LABORATORY TEST COSTS ARE COMPUTED
[157] ⍝   FOR EACH TREATMENT COMBINATION TYPE OF CIGARETTE
[158] ⍝   AND STORED IN COLUMN 9 OF THE MATRIX "DMAT".
[159] ⍝
[160] P9:DMAT[;9]←0.01×⌊0.5+100×((10×DMAT[;1]≠0)+0.5×DMAT[;1])+(0,5,10,20)[(DMAT[;3]÷0.15)+(60×DMAT[;6]-0.07)+(5×~]×
      DMAT[;4])+((0 0 0 30 20 10)[DMAT[;5]]+23)×1+1.2144×DMAT[;2]-0.75
[161] 'TYPE IN YOUR NAME:'
[162] LOC2←⍞
[163] 'TYPE IN YOUR I.D. NUMBER THEN ADVANCE PAPER TO A FRESH
PAGE AND PRESS THE RETURN BUTTON.

NOTE:    RESPONSES ARE STORED IN:  ';,RESP/[1]4 9⍴'       CO       TAR NICOTINE       HCN'
[164] ⍝
[165] ⍝   TREATMENT COMBINATIONS,COSTS AND RESPONSES ARE PRINTED.
[166] ⍝
[167] LOC4←⎕
[168] LAB←12 5⍴' JAN. FEB. MAR. APR.  MAY JUNE JULY AUG.SEPT. OCT. NOV. DEC.'
[169] '


                       RUTHLESS TOBACCO COMPANY
                       ************************


[170] '                                                    INVOICE NO.: ';LOC1
      LOC1←?25                                                  DATE: ';LAB[⌊LOC1÷10000;];' ';⌊(10000|LOC1)÷100;', 19';10
[171] 'TO: RESEARCH BRANCH ';LOC4;'
ATTENTION: ';LOC2
[172] '

                                                              COSTS
ADDI¨   DIA¨    EFFI¨   OVER¨   POR¨            PORT  LOT    OF TEST'
[173] 'TIVE    METER   CIENCY  WRAP   OSITY  WEIGHT   NO.   NO.    SAMPLES ',,RESP/[1]4 9⍴'   CO(MG)   TAR(MG)   NIC(MG)   HCN
      (UG)'
[174] LAB←'F8.3,B16,B16,X3,B5B,F6.2,',(,RESP/[1]4 5⍴'F8.2,F9.2,F8.3,F9.1,'),'X1'
[175] J←0
[176] LOC1←3 5⍴'    S    P   VP'
[177] LOC2←6 7⍴'    C10    C20    C30    P10    P20    P30'
[178] LOC3←6++/RESP
[179] LOC4←+/DMAT[;9]
[180] DESMAT←14↑0
[181] LOOP:' '
[182] I←1
[183] LOOP1:⎕←'F6.2,F8.3,F7.2,X1 'DESMAT[I 3⍴DMAT[I;1 2 3]];LOC1[DMAT[I;4];];LOC2[DMAT[I;5];];,LAB DESMAT[1,LOC3]⍴56
      DMAT[I;]
[184] →((NOPR←1)≥I←I+1)/LOOP1
[185] →(NRUNS≥J←J+1)/DMATDRP
[186] →LCOST
[187] DMATDRP:DMAT←(NOPR,0)↓DMAT
[188] →LOOP
[189] LCOST:LOC1←0.01×⌊0.5+100×(0.2+0.8×NRUNS)×500++/RESP/300 0 300 500
                                                              --------
[190] '
[191] '                         TOTAL COST OF TEST SAMPLES:     $';,'F9.2'DESMAT LOC4
[192] '                         LABORATORY CHARGE FOR TESTING: $';,'F9.2'DESMAT LOC1
[193] '                                                        ----------
                     TOTAL COST OF EXPERIMENT:       $';,'F9.2'DESMAT(LOC1+LOC4);,'
                           ( ';NOBS;' DETERMINATIONS)'
[194] LERASE:⍎')ERASE  DESMAT GLOBVARS'
```

BEST COPY

E: PROGRAM END

```
    ∇ END
[1]   ⍎')ERASE SIMULATE'
[2]   ⍎')SAVE'
    ∇
```

F: PROGRAM DESGEN

```
    ∇ DESGEN;LOC1;LOC2;LOC3;J;LOC4
[1]   'YOU MUST SPECIFY LEVELS FOR EACH OF THE FOLLOWING:'
[2]   LOC1←2 6 12ρ'  ADDITIVE    DIAMETER  EFFICIENCY    OVERWRAP     POROSITY     WEIGHT       [0,0.8]      [.75,.78]
  [.15,.60]      [1,3]        [1,6]     [.97,1.02]'
[3]   'FACTOR:';,((LOC2←FACSTUD[16])/[1]LOC1[1;;])
[4]   'RANGE: ';,(LOC2/[1]LOC1[2;;])
[5]   '
     ALL OTHER FACTOR LEVELS FOR EACH TREATMENT COMBINATION HAVE BEEN
PRE¨SET AS INDICATED PREVIOUSLY.'
[6]   →(0=+/FACSTUD[7 8])/L1
[7]   '
     IN ADDITION, YOU MUST PROVIDE THE FOLLOWING:'
[8]   →(~FACSTUD[7])/L11
[9]   '
        A NUMBER BETWEEN *** 1 AND ';FACLEV[7];'   *** WHICH SPECIFIES
        THE PORT AT WHICH THAT PARTICULAR CIGARETTE WILL BE SMOKED.'
[10]  L11:→(~FACSTUD[8])/L1
[11]  '                                    ';(2 3ρ'   AND')[+/FACSTUD[7 8];];'
        A NUMBER BETWEEN *** 1 AND ';FACLEV[8];'   *** WHICH SPECIFIES
        THE LOT FROM WHICH THAT CIGARETTE WAS SELECTED.'
[12]  L1:'
     ENTER THE TOTAL NUMBER OF OBSERVATIONS REQUIRED. THIS NUMBER
MUST NOT EXCEED ***   ';(×/SIZERES);'   ***.'
[13]  LOC1←⎕
[14]  DESIGN←(LOC1,LOC3←+/FACSTUD)ρ0
[15]  LOC2←+/LOC2
[16]  '
     ENTER THE TOTAL NUMBER OF RUNS REQUIRED. THIS MUST NOT
EXCEED ***   ';SIZERES[2];'   *** RUNS.
     '
[17]  LOC4←LOC1÷⎕
[18]  L2:'
     YOU MAY NOW INPUT YOUR DESIGN, ONE TREATMENT COMBINATION AT A TIME.
FOR EACH TREATMENT COMBINATION, SPECIFY THE LEVELS OF THE ***   ';LOC2;' ***
FACTORS REQUIRED IN THE ORDER LISTED ABOVE (I.E., ALPHABETICALLY).'
[19]  ,FACSTUD[7 8]/[1]2 29ρ'THEN, GIVE THE PORT NUMBER.  FINALLY, GIVE THE LOT NUMBER.'
[20]  '
     START WITH THE FIRST TREATMENT COMBINATION IN THE FIRST RUN
AND ENTER THE REQUIRED VALUES AS A VECTOR OF LENGTH ';LOC3;',LEAVING
A BLANK BETWEEN EACH NUMBER.
     '
[21]  LOC1←LOC1-1
[22]  J←0
[23]  R1:'
SPECIFICATIONS FOR:   RUN *** ';(⌈(1+J)÷LOC4);' *** ; TREATMENT COMBINATION *** ';(1+LOC4|J);' ***'
[24]  →(LOC3=ρLOC2←⎕)/R2
[25]  '
*** INPUT ERROR IN NUMBER OF LEVELS¨¨¨TRY AGAIN ***
     '
[26]  →R1
[27]  R2:DESIGN[J+1;]←LOC2
[28]  →(LOC1≥J←J+1)/R1
[29]  '

     YOUR DESIGN SPECIFICATIONS ARE NOW STORED IN THE ';(1+LOC1);' BY ';(+/FACSTUD);'
MATRIX ¨DESIGN¨.  DO YOU WISH TO CHECK THE DESIGN MATRIX BEFORE CONTINUING
WITH YOUR SIMULATION?   (YES OR NO)'
[30]  →('N'∈⍞)/DCONT
[31]  '



'
[32]  ⍎')DIGITS 6'
[33]  DESIGN
[34]  ⍎')DIGITS 10'
[35]  '


DO YOU WISH TO MAKE CHANGES IN THE MATRIX ¨DESIGN¨?   (YES OR NO)'
[36]  →('N'∈⍞)/DCONT
[37]  LCH:'SPECIFY CHANGE AS FOLLOWS: ROW,COLUMN,NEW VALUE'
[38]  DESIGN[LOC1[1];LOC1[2]]←(LOC1←⎕)[3]
[39]  FCHG:'DO YOU WISH TO MAKE ADDITIONAL CHANGES?  (YES OR NO)'
[40]  →('N'∈⍞)/DCONT1
[41]  →LCH
[42]  DCONT1:'DO YOU WISH TO RE¨CHECK ¨DESIGN¨?   (YES OR NO)'
[43]  →('N'∈⍞)/DCONT
[44]  '


'
[45]  ⍎')DIGITS 6'
[46]  DESIGN
[47]  ⍎')DIGITS 10'
[48]  '


'
[49]  →FCHG
[50]  DCONT:DESMAT←DESIGN
    ∇
```

BEST COPY AVAILABLE

TEACHING EXPERIMENTAL DESIGN: PRESCRIBED BY A MEDICAL STATISTICIAN

Sheila M Gore, MRC Biostatistics Unit, Cambridge

INTRODUCTION

The saying : those who can do, those who can't teach, and those who can't teach write about it inclines me to "had my wheesit" to save face. Only the radical papers by Dr. Mead and Dr. Preece inspire me foolhardily to write a partial prescription for the teaching of experimental design which would have a high compliance amongst medical statisticians. The course I prescribe has been tested on statistics honours students at Aberdeen University and aspects of it, such as practical methods of randomization and teaching by real examples, have found favour with doctors. Adequate trial size to guarantee a fair test of new medicines is a central issue in clinical studies and is discussed in the section 'Adequate Trial Size'. The collaborative nature of much applied statistical work means that the statistician must find a common language with colleagues from other disciplines - clinical trials for example, concern patients, not experimental units. The section 'Collaboration and Consultancy' considers how to give students practice in statistical consultancy. Unreliable data do not warrant the investment of statistical effort; undergraduate statisticians should learn data scrutiny techniques. These are discussed in the section 'Reliable Data'.

PRACTICAL RANDOMIZATION

Undergraduate statisticians should encounter methods of randomization which are practical in the context of (1) sequential patient entry (2) multi-centre trials and (3) ingenious investigators who try to discern the randomization "pattern" by fair means and foul (for example, keeping track of block length or resealing envelopes). Balanced block randomization with fixed small block length is insufficiently secure when treatments are identifiable (Gore and Altman, 1981). Simple randomization has no place except in the largest clinical trials because it appears careless when an investigator ends up with grossly unequal treatment numbers if he intended a randomization ratio of 1:1. Stratification by more than one or two major covariates risks there being many strata which are represented by few patients.

Sequential patient entry cannot be manipulated to satisfy elegant, balanced designs of the type that traditional teaching delights in. Instead medical statisticians use regression models in clinical trial analysis, allowing for most covariates only retrospectively, a practice which is at odds with the rigour of classic experimental design.

As a compromize solution there is a practical method of randomizing patients which minimizes marginal treatment imbalance over several covariates. The method is called minimization (see Pocock, 1979, White and Freedman 1978) and can be programmed easily or kept track of using card indices. Minimization has the advantage of defeating the ingenious investigator. The following is an example of what happens:- a doctor telephones the central randomization office to register his trial patient. Eligibility criteria are checked and relevant covariate information is given - for example : tumour size ≥ 5cms, premenopausal patient, teratment centre Cambridge. These data are keyed in. The minimization routine looks up the marginal treatment totals so far for (1) all Cambridge patients (2) premenopausal women and (3) patients with tumour size ≥ 5cms and works out the treatment allocation which leads to least imbalance, summed over the marginal treatment distributions for the three covariates. In fact the routine will display a ranking of allocations from the treatment which induces least imbalance to the assignment which most exaggerates the imbalance. A probability distribution is associated with this treatment ranking and favours the treatment assignment leading to least imbalance. For example, in a trial which compares early intervention after myocardial infarction by intravenous beta-blockade with no intervention, the probability distribution might be chosen so that the chance is 2/3 that the patient is allocated to the treatment which reduces the marginal imbalance between treatment numbers.

REAL EXAMPLES

Even trivial designs (as in clinical trials) are challenging in application. Consider a within-patient study of dinitrochlorobenzene (DNCB) for the treatment of alopecia areata, a condition which entails partial or complete hair loss. DNCB dissolved in acetone was painted onto one half of the patient's scalp, the other half being untreated. Even discounting the punk appearance (in 1977!) of those patients for whom DNCB induced effective hair growth, the design has two other limitations. Firstly, the action of DNCB may be systemic as well as topical and so the untreated side may benefit from a systemic action; secondly there is contamination of the untreated area when the patient rubs or scratches his head. Choosing an appropriate study design is one of the most challenging aspects of problems in applied statistics and so experimental design must surely be taught through real examples which are not "shorn" of context.

Curiosity is a help to the applied statistician in discovering the context of a problem:

"I have six honest serving men,
They taught me all I knew,
Their names are what and why and when,
And how and where and who".

Rudyard Kipling

ADEQUATE TRIAL SIZE

A fundamental problem in the design of clinical experiments is one to which neither Dr. Preece nor Dr. Mead accorded much speculation. It is patient accrual or adequate trial size. In a survey of 38 clinical trials in The Lancet (July-December 1977) I found that no study mentioned how trial size was determined. More recently Dersimonian et al. (1982) reported on 67 trials published in four leading general medical journals. Power considerations were omitted from 85% of the reports. Pocock (1978), on the other hand, surveyed trial organisers and found that in a majority of randomized cancer studies statistical methods were used to determine trial size. Not until these statistical aspects are routinely written in to published reports will the ideas become familiar to doctors. Differences which are not statistically significant in trials with inadequate accrual are often wrongly interpreted (Freiman et al. 1979) as meaning that there is no difference between treatments. Were the corresponding confidence interval reported, it would be seen to be hopelessly wide and a ridiculous basis for deciding between treatments. I regard studies which have less than 50% power to establish a worthwhile treatment difference as bad science. Dr. Preece points out that there is virtue in combining the results from such trials. Indeed there is, provided that one avoids selection bias. Because of the tendency for unpromising studies to fall by the wayside and go unreported, there is an inherent bias in combining the results from published trials only. The pharmaceutical industry is the best place to keep and make available a register of all clinical trials for which it releases supplies of new medicines.

Therapeutic advances - such as the use of anticoagulants (Chalmers et al. 1977) and intervention by beta-blockade after myocardial infarction - have been delayed by inadequate trial size and failure to report confidence intervals. It is perhaps more than mere coincidence that the first trial of beta-blockade which showed a significant reduction in mortality after myocardial infarction (Norwegian Multicenter Study Group 1981) was also the first which was large enough to detect a worthwhile treatment difference. "Not statistically significant" does not mean "there is no difference". Too often it means still: "there has not been a fair test of this treatment"!

That the teaching of experimental design does not emphasise enough how to determine patient numbers was evident at a recent SERC conference in Reading for postgraduate students in statistics,few of whom knew how to calculate trial size. Is it any wonder that doctors get it wrong sometimes? An incidental advantgae of the promotion of stopping rules and group sequential methods (Pocock 1982) for clinical studies is that they impose an explicit discipline on trial size.

COLLABORATION AND CONSULTANCY

When I gave an honours course on clinical trials in the Department of Statistics at Aberdeen I invited clinical colleagues to attend. The presence of a doctor in the lecture room was valuable support for me when explaining the medical background to problems and was a reminder to students of the collaborative nature of much applied statistical work. Of course, I gave notice of the more technical lectures which doctors would not find it profitable to attend but always included on their return a resume of the essence of the statistical method we had been studying. Thus I encouraged students to recognise that what is important for a medical audience is being able to understand the idea behind the formal mathematics, and that communicating such ideas is part of their responsibility as statisticians. Likewise, my recommendation about interpreting and reporting results, especially in relation to complex data, was for the use of graphical methods and descriptive statistics to record initial impressions. These are the stepping-stones by which the statistician reaches a final analysis and are no less important for the reader if he is to have some feel for the data and understand better the results of statistical modelling.

Even before report-writing, students must develop a skill for statistical consultancy. I have learned that divulging to students the steps which I take to become conversant with a medical problem before the first consultation and the strategems I have for verifying what a doctor tells me is no substitute for their practising this for themselves. When lecturing on clinical trials I therefore posed to each honours student (numbers varied from three to seven students) a separate clinical trial problem, as it had been brought to me. The students arranged a meeting with me, who played the part of the clinician, and they got only such additional information about the problem as they thought to ask for. This is precisely what happens in practice: if one fails to ask a relevant question, the information is not forthcoming as the doctor may not appreciate its statistical value. Several students said how relieved they were that their first experience of consultation was "in house" because their fund of questions was too soon exhausted. Only then did they take seriously the points I had ventured about being prepared for a consultation! Moreover, many found it necessary to request further meetings when they got down to writing the protocol for the clinical trial because they became aware of salient gaps in their knowledge of the problem.

I suspect that too often, when lecturers talk about practical classes and practical exercises, what they mean is practice at routine design and analysis in relation to some rounded problem. But it is the ragged edges that statistical consultancy is all about. Discovering these and weighing them in the balance is why applied problems are seldom routine, and are never dull.

RELIABLE DATA

At the end of a lecture on the checking and editing of medical data I was surprised to be asked "Who has the time to do all that?". My reply was "Who can afford not to do it!". Students should learn about and have practice in data scrutiny. No amount of analysis can make bad data good.

Transcription errors may be checked by reading across or by double-punching; they can be avoided by keying in data directly to a microcomputer which has a well designed and tested screen format. Range checks and checks on the accuracy of recording are elementary and can be programmed, as can logical checks: for example it is inconsistent that a patient who is 60 years old at diagnosis should be menstruating regularly, impossible that date of treatment is later than date of death and so on ... Arithmetic errors can be avoided by programming the relevant calculations - in two separate studies I have detected 15% and 40% respectively error rates for age last birthday. Sorting data by trial number and/or date of randomization, date of birth may detect duplicate patients or missing serial numbers. Even after all such checks have been made there may be residual errors which have transgressed neither logical nor range checks. To estimate the residual error rate, one selects a random sample of patients and compares their computer records with hospital case notes. I did such an exercise for the Western General breast cancer series and discovered an important coding anomaly which had been overlooked hitherto. Of the 361 patients whose tumours were ulcerated, fixation to the overlying skin was recorded as nil or complete for all but 9 patients, the extremes being noted with similar frequency. The explanation was that no coding convention had been determined, and the consultant had argued variously:

(i) the tumour is ulcerated and so has broken through the skin, having been previously completely fixed to it. Code: complete fixation.

(2) ulceration supersedes fixation, which is thus irrelevant. Code: nil fixation.

A salutary tale on which to end.

References

[1] Dersimonia, R., Charette, L.J., McPeek, B. and Mosteller, F. (1982) Reporting on methods in clinical trials. N.Engl.J.Med, 306, 1332-7.

[2] Freiman, J.A., Chalmers, T.C., Smith, H. and Kuebler, R.R. (1978) The importance of beta, the type II error and sample size in the design and interpretation of the randomized control trial: survey of 71 "negative" trials. N.Engl.J.Med, 299, 690-4.

[3] Gore, S.M. and Altman, D.G. (1982) Statistics in Practice, British Medical Association, London, 44-47.

[4] Norwegian Multicenter Study Group (1981) Timolol-induced reduction in mortality and reinfarction in patients surviving acute myocardial infarction. N.Engl.J.Med., 304, 801-7.

[5] White, S.J. and Freedman, L.S. (1978) Allocation of patients to treatment groups in a controlled clinical trial. Br.J.Cancer, 37, 849-857.

[6] Pocock, S.J. (1978) Size of cancer clinical trials and stopping rules. Br.J.Cancer. 757-66.

[7] Pocock, S.J. (1979) Allocation of patients to treatment in clinical trials. Biometrics, 35, 183-97.

[8] Pocock, S.J. (1982) Interim analyses for randomized clinical trials. Biometrics, 38, 153-62.

CONTRIBUTION TO DISCUSSION OF THE ABOVE PAPERS

B Jones, University of Kent

Dr. Preece has indeed told us "what it is about", and drawn our attention to examples of some of the misconceptions and careless thinking which has crept into some published accounts of designed experiments. However, an important question I would like to put to him is "How to teach such a large subject in a typical university lecture course?". In Table 1 I have given a division of Dr. Preece's "top 12" components. The upper half includes what I consider tends to get taught, certainly to students of mathematics, to those studying the design of experiments. The lower half contains those topics which I feel are better placed in a separate and later course which aims to cover the practical application of the theory. I regard both halves of the table to be of equal importance but such a subdivision makes it more convenient from the teaching point of view. A typical lecture course consists of 30-40 hours of lectures and so each half could be accommodated in a standard lecture program.

TABLE 1: A division of DAP's "Top Twelve" headings

Theory

1. The basic principles of statistical inference.
4. The basic principles of "good" design.
5. Mathematical methods for constructing designs and obtaining analyses, particularly analyses of variance and covariance.
9. Ability to recognise and interpret patterns and anomalies in data.
10. Understanding of the assumptions underlying standard statistical analyses, and the ability to recognise when these assumptions are seriously violated in practice.
11. Facility in spotting mistakes in analyses.
12. Skill in reporting conclusions in clear, simple prose, with appropriate tabulations and graphs.

Practice

2. The ability to select treatments, experimental units, and variates that will permit "useful" results to be obtained.
3. The ability to recognise relevant sources of variability, and to make preliminary assessments of their importance and relative magnitudes.
6. Trained personnel for setting up and managing experiments.
7. Proper instruments (balances, etc.) for obtaining accurate, precise data values, and trained personnel to use the equipment.
8. Training in data transcription (with checking).

Arranging for students to gain real experience of the practical side of the subject is difficult. Apart from a few visits to an industry or research station to see real experiments in progress I think most practical skills can be learned by carefully studying case histories. The particular university course I teach is quite short and so I am able only to give a flavour of the practical side. This I do by making use of a large collection of projector slides of photographs of real experiments. These are used to illustrate what experiments look like, the size and shape of plots and the details recorded on the working plan of the design e.g. sites of former roads or streams in an agricultural trial. Spending a lot of time visiting one real experiment is usually not justified because a student only gets exposed to a limited point of view. The student certainly gets a feel for the problems involved but only gets experience of one experimental situation. For example, studying the practical difficulties of designing a raspberry trial may not be much use to a student who is eventually employed to design trials on cowpea in a developing country.

Another question I would like to put concerns the approach we ought to adopt when teaching the theory. Should we present the subject by way of numerous special cases of interest or should we start with a general theory and then begin to specialize? Having tried both I am clear that they each have certain advantages and disadvantages.

With reference to the "three R's" of Fisherian experimental design it seems to me that the one corresponding to randomization stands out as being far more difficult to teach than the other two. All three are quite intuitively reasonable but the large theoretical framework associated with randomization is often daunting to most students.

Mr. Mead's ideas on using computers as an aid to teaching are very useful. The use of simulation games provides a worthwhile alternative to lengthy visits to experimental sites. I have since tried his Tomato game with an M.Sc. in Statistics class and they found it enjoyable and instructive.

His overall approach is attractive: we can now let computers take the burden of analysis so let's concentrate on getting the most appropriate design. Often in the past experiments have been designed a particular way because the analysis was available in a textbook, without any regard for whether or not the most suitable experimental design was used. The use of some partially balanced incomplete block designs certainly could be criticised from this point of view. The use of a computer program such as GLIM is definitely required to make the most of Mr. Mead's ideas. However, teaching students to learn GLIM just for a design course, would probably not be practical. Therefore, if such a program is to be used it should have already been taught as part of the overall degree course in Statistics.

His lack of emphasis on the usefulness of factorial experiments is regrettable as these are widely used, particularly in industry. However, his suggestion that we should concentrate on getting desirable patterns of treatment concurrences is most important but has received very little obvious attention in standard text books. In a factorial experiment there are usually some effects, e.g. high order interactions which are of little interest, while others e.g. main effects must be well estimated. Finding patterns of treatment concurrences which lead to designs where the most important effects are well estimated is often not easy but provides a useful basis for introducing students to the idea that standard text book designs may be of little use on some occasions. Also it helps introduce another feature of the computer - that of being programmed to search for 'optimal' designs. To illustrate this point of view I have used an algorithm (Jones and Eccleston, 1980) to search for designs in 6 blocks of 4 plots for a 2 x 3 factorial experiment. The two factors I have labelled as L and N. Factor L has two levels A and B and N has 3 levels 1, 2 and 3. A set of five possible contrasts of interest is given in Table 2, where each contrast is scaled so that the sum of squares for its coefficients is unity. Table 3 gives the variances of the five contrasts obtained from separate runs of the algorithm. In each run a degree of importance or weight was assigned to each contrast, where a weight of 1 indicates full importance and O indicates no importance. The algorithm requires a starting design and the one used in each run was that given by Mr. Mead in his Figure 2. It will be noted that Mr. Mead's design cannot be improved upon if all contrasts are of equal importance, i.e. the weights are (1, 1, 1, 1, 1). It is clear that there is a wide choice of design depending on the experimenter's interests. The algorithm used searches for optimal concurrences of treatments and so is fully in the spirit of Mr. Mead's suggestions.

TABLE 2: Treatment Contrasts

	A1	A2	A3	B1	B2	B3	Scaling value	
1	1	1	1	-1	-1	-1	$^1/\sqrt{6}$	MAIN EFFECTS
2	1	0	-1	1	0	-1	$^1/\sqrt{4}$	
3	1	-2	1	1	-2	1	$^1/\sqrt{12}$	
4	1	0	-1	-1	0	1	$^1/\sqrt{4}$	INTER-ACTIONS
5	1	-2	1	-1	2	-1	$^1/\sqrt{12}$	

TABLE 3: Jones and Eccleston (1980) Designs obtained from Roger Mead's Starting Design

Contrast variances

WEIGHTS	1	2	3	4	5
1 1 1 1 1	0.30	0.27	0.27	0.31	0.26
1 1 1 $\frac{3}{4}$ $\frac{3}{4}$	0.25	0.27	0.27	0.31	0.31
1 1 1 $\frac{1}{2}$ $\frac{1}{2}$	0.25	0.27	0.27	0.31	0.31
1 1 1 $\frac{1}{4}$ $\frac{1}{4}$	0.25	0.27	0.27	0.31	0.31
1 1 1 0 0	0.25	0.27	0.27	0.8	0.31
$\frac{3}{4}$ $\frac{3}{4}$ $\frac{3}{4}$ 1 1	0.30	0.31	0.27	0.27	0.26
$\frac{1}{2}$ $\frac{1}{2}$ $\frac{1}{2}$ 1 1	0.27	0.32	0.30	0.26	0.25
$\frac{1}{4}$ $\frac{1}{4}$ $\frac{1}{4}$ 1 1	0.25	0.33	0.33	0.25	0.25
0 0 0 1 1	0.25	0.33	0.33	0.25	0.25

A point I find puzzling however is his suggestion that only unweighted concurrences should be considered. I feel it is only the weighted concurrences which matter and give in Tables 4 and 5 some examples which support this view. In each example the experimenter is likely to be misled by looking at the unweighted concurrences.

In conclusion I congratulate both speakers on making plain the fact that the subject of experimental design is full of interest and is able to take advantage of the advent of computers.

Reference

[1] Jones, B. and Eccleston, J.A. (1980) Exchange and interchange procedures to search for optimal designs. J.R. Statist. Soc., B, 42, 238-243.

TABLE 4: Example of Unequal concurrences and Unequal Block Sizes 6 Treatment in 18 Blocks - Block Sizes 6 x 4, 12 x 2.

BLOCK				
I	1	2	3	4
II	[illegible]	2	5	6
III	[illegible]	4	5	6
IV	1	2	3	4
V	1	2	5	6
VI	3	4	5	6
VII	1	3		
VIII	1	4		
IX	1	5		
X	1	6		
XI	2	3		
XII	2	4		
XIII	2	5		
XIV	2	6		
XV	3	5		
XVI	3	6		
XVII	4	5		
XVIII	4	6		

Note: In this design there are 4 concurrences of (1,2), (3,4) and (5,6) and 3 concurrences of all other pairs.

If only unweighted concurrences are considered we should expect treatment differences involving (1,2), (3,4) and (5,6) to have smaller variances than the rest.

However this design is totally balanced - the sum of the weighted concurrences of every pair is equal.

In particular every standardised treatment difference has variance $\sigma^2/6$.

TABLE 5: Example of equal concurrences and unequal block sizes

BLOCK				
I	1	2	3	4
II	1	2	5	6
III	3	4	5	6
IV	1	3		
V	1	4		
VI	1	5		
VII	1	6		
VIII	2	3		
IX	2	4		
X	2	5		
XI	2	6		
XII	3	5		
XIII	3	6		
XIV	4	5		
XV	4	6		

Notes: Here the unweighted concurrences are equal but not all treatment differences are estimated with the same precision.

E.g. Variance of 1 vs 2 is $0.25\sigma^2$
Variance of 1 vs 3 is $0.23611\sigma^2$.

Indeed the design is group divisible - so again the weighted concurrences are the only things of interest.

TOPIC 14: TRAINING INDUSTRIAL PRACTITIONERS

ORGANISER: DR R D SNEE

THE POLYMATH CONSULTANT

A A Greenfield, The Queen's University of Belfast

INTRODUCTION

I speak from experience as a consultant. This experience was garnered mainly in two very different industries: the steel industry and the medical faculty of a university. It is supported by a brief consulting experience in the pharmaceutical industry, by employment in quite different roles in the mining and publishing industries, and by brief contacts with the gas, oil, chemical and quarrying industries. It was through all these diverse experiences that I eventually reached two conclusions: one is that there is a profitable role in life for a polymath; the second is that a man or a woman can be taught to be a polymath.

After I had reached these conclusions, I made a very pleasing discovery. This was that my friends at Sheffield City Polytechnic, Kanji and Gilchrist (1981), had reached the same conclusions by an entirely different route. But instead of the word 'polymath' they use the term 'applied statistician'. Furthermore, they are doing something about it: they are running the best course, in terms of practical value, that I have been able to find anywhere.

A polymath is a person who knows many arts and sciences. In an earlier paper (Greenfield, 1979) I described the role of the polymath in industry, the great benefits he would bring to it, and the pleasure he would derive from such employment. In this paper, I shall briefly restate those points and then proceed to propose the outline curriculum of a three year undergraduate course leading to a Bachelor of Science. I shall leave it to you to imagine the fine detail of the course and hope that I shall stimulate you to add some constructive suggestions.

Since writing that paper, I was greatly encouraged to read a report by the American Statistical Association's committee on Training of Statisticians for Industry, entitled "Preparing Statisticians for Careers in Industry". In this, they acknowledged a close agreement with my ideas and they discussed the major approach to their task. However, even they did not go far enough in describing the best possible education for an industrial consultant. Even though the broad education they propose, and Sheffield City Polytechnic and others provide, will produce excellent applied statisticians, and even though I agree that statistics is essential to my proposal, there is a need in industry at large for a brighter jewel; the applied polymath. And I believe that he could be produced, post A-level, in three hard years of study and active practical learning.

WHO NEEDS POLYMATHS?

British industry as a whole needs polymaths because Britain lacks many of the natural resources that other countries possess and we depend for our prosperity on manufactured products. There has been a widespread failure in recent years to put more resources into product development. This has not usually been for lack of finance, nor for lack of government investment incentives. Nor is it for the lack of advice on how to do it, because academics and management consultants are for ever telling the industrial managers how to do it. I believe it is because management structures have become trees of little boxes, each filled with a specialist who knows his own place. Creativity has been cramped. The way to overcome this is to employ super-statisticians. Not super in the sense of theoretical and academic brilliance, but super in the sense of knowing something, preferably a lot, about most things that make industry tick: the sciences and technologies, people and markets, accountancy, finance, economics and business systems, and, above all, communications because it is communication that cuts across all those barriers.

But why a statistician? Why not a super-something-else? Professor Michael Healy (1979) gave the answer in his inaugural lecture:
"The statistician can contribute technological suggestions about the conduct of the research which may greatly increase the impact or reduce the cost; but he can contribute a great deal more than this if he has an adequate broad knowledge of the subject matter. He is the one who is most likely to spot the analogies between the problem in hand and other problems elsewhere in medicine, in agriculture or in industry. He may be the only one to suspect that such an analogy exists. He will also be the least likely to carry such analogies across uncritically into the new area and the most capable of adapting them to the new requirements".

Broadness of view and ability to see the big picture are needed in any dynamic and enterprising business, because of the diversity of problems and situations that will be met. It troubles me that so many statistical studies are confined: statisticians restrict themselves to the experimental design and analysis in a formal quasi-academic cage without allowing themselves to reach out into the wider issues which they could so easily influence if they tried. Hunter (1981) argues similarly from his own experience.

SOME EXAMPLES

In my earlier paper (1979) I briefly described the structure of the steel industry and explained one process, steelmaking, in more detail to illustrate the need for a useful statistician to be able to understand, communicate with, and help such diverse specialists as: physicists, chemists, metallurgists, psychologists, accountants and engineers of all sorts: electronics, chemical, mechanical and production.

Figure 1 shows that in the pharmaceutical industry there is a similar diversity of links. This is contrary to popular opinion.

When I asked the academics in an English university: "What should be the academic training of statisticians in the pharmaceutical industry?" they all gave me the same answer. I asked in the statistics department. "Clinical trials", they told me. I asked in the pharmacology department. "Clinical trials", they said. And in the departments of physiology, biochemistry and community medicine, they all told me: "Clinical trials".

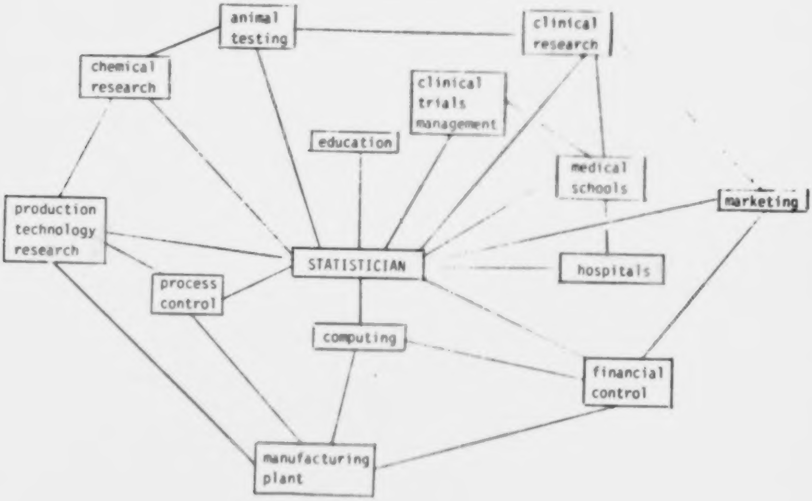


FIGURE 1: The Statistician in the Pharmaceutical Industry

In examining the problem, I have tried to imagine the life of a statistical consultant in the industry with a view to drawing the links between his life and his needs for academic training. But in thinking about his life, I found myself dwelling more on his desirable characteristics, his background and his interests rather than what is currently regarded as formal curriculum studies. I hope to identify those qualities by describing the possible role of the statistician in the manufacturing, commercial and research environment. My example is based on a brief attachment to a pharmaceutical company although the conclusions relate to all other industries based on the physical, chemical, biological, or engineering sciences.

Opportunities for, and worth of statisticians in all industries are far from fully appreciated. The well-trained statistician is more suited than any other specialist to take a leadership role in the solution of company problems, whether in research, development, manufacturing, commercial or personnel areas. But what is well-trained?

The well-trained statistician is the one with personality and panache, with an understanding of company structure and operation, with a wide based knowledge of science and engineering, and with an ability to communicate. Each of these is as important as a thorough grasp of statistical theory and methodology.

Just look at the structure of the pharmaceutical industry. Many companies are now international. Some have more than 40 peripheral associate or subsidiary companies distributed across the face of the globe constituting the nodes of a network of manufacturing, research and commercial activities. The statistician must be prepared to work in such an organisation and contribute generally to its success.

Think of a single unit within such a great complex: one research laboratory. This laboratory is manned largely by scientists and engineers who have been educated deeply and narrowly in special fields, often to the doctorate level, plus associated technicians. There are medics, psychologists, biochemists, botanists, zoologists, and maybe plant engineers and control engineers. The roles of these specialists are to invent, develop and test new products and processes. These aspects of research and development are, in my view, the most exciting, the most intellectually appealing, and the most rewarding. The rewards are in the pleasure of personal achievement, in contributions to scientific progress, and in enhancing industrial profitability. But these specialists should also be in touch with their counterparts in industrial production (on the works) and in commerce (in the marketing companies). Here their role, associated with plant in use and with current products, is to identify problems affecting the efficiency, and therefore the costs, of working and marketing practice, and to solve these problems. These may be associated with energy, manpower, health and safety, consumable materials, product quality, test procedures, plant maintenance and life, packaging, distribution, storage, and legal and ethical conditions related to the products.

Within that organisation, let us identify what a statistician actually does, or what his colleagues expect him to do, rather than what his job description says he does. Think too, about what a statistician could do, to the benefit of himself and of his employer, IF he had the somewhat broader training that I shall advocate. Furthermore, consider what he would do if that training included practical instruction on how to assert himself.

All of you with experience in similar organisations, know that the statistician has daily to practice the skills of consulting, or being consulted; he is involved in operations research, in simulation, in computing, in running courses, in chairing committees, in speaking in public, and in writing reports. His involvement, for example, in clinical trials is not restricted to the design of the trial and the analysis of the data. He must advise the medical specialists on what is practical as well as scientific and ethical and he must press them to keep to the timetable and to the protocol. He must set up and monitor management procedures for processing records and he must ensure that there is the proper flow of material in and out of the computing department. He must communicate with, and educate, the staff in peripheral companies and participating hospitals, and he must ensure that adequate supplies of materials and forms are reaching the area manager and being forwarded by him. He must keep the clinical research director informed about the daily progress and report to him all the things that do go wrong with big trials with suggestions as to how to put them right. Even without describing any projects in detail, you will appreciate that the statistician will be expected to understand and help medics, scientists, and business men, even though all these specialists will arrogantly remain ignorant of the statistician's arts.

Thus the mathematical or statistical consultant must have a wide scientific and technical education if he is to succeed in industrial research. That education must include written and spoken communication with scientists and engineers of different disciplines. Such polymaths already exist, but usually their wide abilities have developed over a decade or more after a more specialised formal edcuation. That formal education should and could be modified to produce more polymaths so that they may be more useful to research and development and the rest of industry earlier in their careers. There is a helplessness and potential bankruptcy in specialisation, but there is security and potential wealth in versatility. Education should aim for that versatility with particular qualities: a thorough grasp of scientific method acquired through formulating and solving practical problems; a broad knowledge of scientific and technical subjects relating to the major industries; an understanding of industrial organisation and management; and, emphatically, a well-developed ability to communicate at all levels and in many different circumstances.

This list of qualities, you may object, are a long way from statistics. I think not. The statistician is central to any situation where data are collected, analysed, interpreted and used. There is no occasion in industry or industrial research where this does not apply. A statistical education is one of two keystones to this degree in polymathy that I envisage. The need for good communication is the second keystone and it is universally agreed. Yet little is done about it after O-levels. It is a sad fact that those students who choose the sciences rarely have any instruction, practice, or correction in communication after the age of 16, when they take English language at O-level. Then, two or three decades later, after they have been employed in industry for several years, either they or their managers realise how thoroughly bad they are at communication. They are sent on brief, superficial, and expensive management courses. Those who employ new graduates are often heard to grumble about the poor articulation, the bad grammar, and the wrong spelling of their recruits.

THE POLYMATH COURSE

An approximation to polymathy in some degree institutions is the modular method by which students may choose modules from the full arts and sciences spectrum. They win points towards their final degrees as they succeed in each module. This does not seem to me to be such a good idea because it does not yield a coherence of approach, knowledge and experience. Chemistry, for example, is distinct from English which is distinct from statistics. On the other hand, my proposal, although it could be seen as a list of many disparate subjects, is planned to be a merging of those subjects into a unified whole. It is a reflection of creativity, which can be defined as an ability to associate those things which are not associated so as to produce something new. Perhaps the best way to explain how this can be done is by describing my draft three year timetable shown in Figure 2.

This timetable is only a first draft. I hope that discussion will lead to improvements. But it will serve for the present to illustrate my intentions.

There are two main features of the course. One of these is what you would normally expect in any undergraduate course, but do not always find: it takes the final school year material, in the sciences, and uses it to illustrate and practice the philosophy of the course before introducing new material and building towards a greater structure of knowledge, experience, understanding and ability. The second feature is that each subject is taught in such a way as to link closely with all other subjects. None of them is a subject on its own. They are all mutually intermeshed. For example, chemistry, writing, and statistics are mutually dependent on each other. Thus, to illustrate both main features, sixth-form scientific laboratory work will be repeated in the first term, but with emphasis on data recording, analysis and presentation, written reports, conference discussion, and touching on the needs for mathematical modelling and experimental design.

COMMUNICATION

We all know well enough that the statistical industrial consultant must be able to interview a client, to persuade him to talk about and help him to formulate his problems. He must be able to write memoranda and letters explaining his work and seeking information and action. He should be able to write reports and articles describing how he has tackled and solved problems. He should be able to speak in public, spontaneously as well as in the formal presentation of papers. He should know how to behave on committees and working groups, as chairman, secretary, or contributor. He should have a working knowledge of making diagrams and charts, slides and transparencies, block and plate making, typography, printing, proof-reading and layout of publications. He should be trained in self-communication: note-taking, transcription, and filing.

Although there are volumes about the theory of communications, the real worth of any course is in the eventual ability of the students. We should therefore aim at the development of practical skills in terms of speaking ability, writing ability, and the ability to use the media of communication. One of the most useful of these abilities is typing and I propose that this should be required in the first term of the course. A further advantage of this skill is its value in using computer terminals and word processors. This should be followed by other physical techniques of communication: printing, photography, graphic design, and visual aids for lectures, displays and exhibitions, including the making of slides, overhead projector foils, projection techniques, models, films, video equipment, and computer developed graphics.

The skills of writing would be developed by formal instruction about the writing and editing of project reports, scientific papers, proposals for research and investment, technical articles, instruction manuals, memoranda, committee minutes, letters for business and newspaper publication, the use of word processing. There would be sessions on the analysis of published papers and reports.

All of these written and mechanical communication abilities would be brought together in the third year when three hours a week would be devoted to the production of two course publications: a weekly newspaper and a monthly magazine. These would contain the products of other parts of the course, which I shall describe later, as well as any other material produced by staff and students.

Public speaking abilities would be developed in the first year together with interviewing skills. Public speaking would include the arts of impromptu speaking, debating, speechmaking, teaching, and public lecturing. Interviewing would be helped by third year students who would act as interviewees, about simulated or real problems, for the first year students. In all these activities, mutual criticism and competition would be encouraged to develop personality and confidence and to overcome faults.

Speaking exercises would continue through into the final year when there would be a weekly debate on current affairs. This would be organised with formal rules of debate and course points would be awarded for participation. Committee work would be simulated and there would be brainstorming sessions to stimulate creativity.

All practical work would be reported by the students, using visual aids, as if they were teaching or delivering papers, and they would be encouraged to criticise each other's presentation as well as content. These presentation and discussion sessions appear prominently on my draft timetable. The weekly climax of these activities would, however, be the course conference, occupying the last one and a half hours of the Friday afternoon. The students of all three years would gather together for this. Although presentations would mainly be by senior students, others would be encouraged to join in.

All of this practical learning would be supported by more formal instruction: in the first year about library and information systems, abstracting, indexing, referencing, researching, and about the theory and art of communication: in the second year about the structure of business and public institutions and the flow and control of information; and in the third year about committee structures, their organisation and operation, about public relations, and about the organisation of courses and meetings.

During the second and third year, each student would study in each term:

1. A paper from one of the more mathematical scientific or statistical journals. He would prepare from this:
    a) an oral presentation, with visual aids, aimed at a general scientific audience.
    b) a written paper, version of (a), for which he would do all the editorial preparation including graphics for production in the monthly journal.
    c) a newspaper account of the work, with explanations of possible social, industrial, or scientific implications.

    d) an interview role, to give the first year students experience of interviewing.

2. A paper from a trade or technical publication, which:
    a) he would criticise constructively, including analysis of data published with the article, or supporting data retrieved from other sources, in a written paper for publication in the monthly journal.
    b) a newspaper account of the work.

PRACTICAL PROJECTS

My proposal for dealing with practical projects illustrates the integrated nature of the course. In most courses, for example, if a student were asked to prepare some ethyl acetate he would be exercised only in his chemical laboratory ability. In this course, however, he would also be exercised in model formulation, the statistical aspects of experimental design and analysis, in data collection and recording, in writing and graphic presentation, and finally in oral presentation to the class the following day.

Such projects would be based first of all on subjects with which the incoming student is already familiar, because they would be drawn from A-level science material in chemistry, physics, biology and mechanics. They would be used further to develop his broad scientific knowledge, so that in the second term of the first year his projects would be to do with psychology, electronics and physiology, while in the third term some industrial projects would be simulated in the laboratory. One afternoon of every week in the second year would be devoted to an environmental project, extending his interests into geology, meteorology, pollution, industry, transport, architecture, housing, education and other public services. All such projects would be designed to give experience in hypothesis statement, modelling, data collection, library research, interviewing, data analysis, reporting, and interpretation, as well as written and spoken presentation. Other project work in the second year would be laboratory based, with computer simulation of industrial and other institutional situations and processes. There would also be data analysis projects.

All these would lead to larger and individual research projects for the students in the third year when there would also be half day attachments each week to other institutions, such as manufacturing industry, commerce, research laboratories, and local government and services. Kanji (1979) has discussed the role of projects in statistical education.

MATHEMATICS, STATISTICS AND COMPUTING

Mathematics is not only a basis for studying and understanding the sciences and other subjects the student will meet, it is an international language, a medium of communication. It should therefore be spread throughout the course, as are the other communications studies, rather than being confined to the early terms. There are some important aspects of mathematics which must be included to support other subjects, particularly probability, statistics, operational research, computing, electronics and instrumentation. While these may continue throughout, the later terms should be devoted more to the analysis and discussion of published mathematical applications papers.

The most important statistical techniques, in my experience, are: graphical display and data summary; design of experiments; regression analysis; response surface methodology including evolutionary operation; discriminant analysis; simulation, both process and event; time series analysis, including the analysis of continuous multivariate processes with feedback; numerical analysis; surveys, reliability; quality control and acceptance sampling. Others have views about an applied statistics curriculum and it is being covered well in other papers so I shall not dwell on it here. However, a vital feature is the statistics laboratory, described by Kanji (1974) which he has shown will add greatly to the student's practical understanding of the subject.

The study of computers must be planned so as to produce not only a good user of computers, but also a good adviser about them. It should cover the full range from hand held machines, through micros, minis and up to the large mainframes. Hardware studies should include electronics and instrumentation because these are so often at the root of data collection in industry. Sadly, I have seen too many statisticians assume that the data presented to them must be correct and they spend time analysing errors which they may be able to eliminate if they understood the plant and transducers which produce peripheral equipment: typewriter terminals, VDUs, printers graphics terminals and plotters. They must learn something about computer packages, interactive statistical analysis, textual storage and retrieval, word processing, computer based learning, and simulation. They should be able to program in several high level languages (Basic, Fortran 77, Pascal and Cobol) as well as in an assembler. They should be familiar with documentation, standards, file handling, and structure of languages and programs. Their practical work should yield fully documented programs, illustrating their mathematical, logical, artistic, creative, technical and communications skills.

GOVERNMENT, INDUSTRY AND BUSINESS

It is surprising that so few courses that prepare students for lives in industry actually equip them with any general knowledge about industry. We should start with a study of government and public institutions, banks and insurance companies, progress through the structures of industry and types of business and the links between government, industry and business. Links with other subjects on the curriculum would be studied through business management, economics, financial planning, accounting, manpower planning, industrial relations, operational research and public relations. This may seem to be rather much, but I do not suggest that these studies, except for operational research, should be in any depth. Rather an understanding of the meanings and relationships of these subjects is intended.

CONCLUSION

Perhaps my proposal for a degree course in polymathy is still sketchy, but if I have succeeded in conveying the intended philosophy we may be able to use this paper as a basis for generating further ideas and a detailed course proposal. The advantages are clear. We should produce graduates who are capable of ferreting out problems at any level in industry, solving them, and communicating the solutions so that they are put to profitable effect. Although our graduates would be good applied statisticians, and better than most because of their communicating abilities and broad understanding of the sciences, they would also be potential leaders of dynamic and enterprising businesses within any industry. Furthermore, our graduates would be educated to enjoy interesting lives and to be paid well for living them.

References

[1] Greenfield, A.A. (1979) Statisticians in industrial research: The role and training of the industrial consultant. The Statistician, 28, 71-82.

[2] Healy, M.J.R. (1979) Does Medical Statistics Exist, BIAS, 6, 2, 137-182.

[3] Hunter, W.G. (1981) Six statistical tales. The Statistician, 30, 107-117.

[4] Kanji, G.K. (1974) The role of the statistical laboratory in the teaching of statistics. Int.J.Math.Educ.Sci.Tech., 5, 53-57.

[5] Kanji, G.K. (1979) The role of projects in statistical education. The Statistician, 28, 19-27.

[6] Kanji, G.K., Gilchrist, W.G. (1981) A degree in applied statistics. The Statistician, 30, 193-198.

[7] Preparing statisticians for careers in industry: Report of the ASA section on Statistical Education Committee on Training Statisticians for Industry (1980). American Statistician, 34, 65-75.

FIGURE 2: Draft Timetable for B.Sc. course in polymath consultancy

FIRST TERM

FIRST YEAR

	Monday	Tuesday	Wednesday	Thursday	Friday
9.00	MATHS	MATHS	MATHS	MATHS	MATHS
10.00	COMPUTING	Library and Information	Library and Information	COMPUTING	Psychology
11.00		Presentation chemistry	Presentation physics		Presentation biology
12.00	Institutions	Elementary Statistics	Elementary Statistics	Public speaking Interviewing	Statistics Laboratory
2.00	Chemistry Laboratory	Physics Laboratory	GAMES AFTERNOON	Biology Laboratory	STATS LAB.
4.00	TYPING	TYPING		TYPING	COURSE CONFERENCE

SECOND YEAR

	Monday	Tuesday	Wednesday	Thursday	Friday
9.00	Experimental Design	Experimental Design	COMPUTING	Experimental Design	Experimental Design
10.00	MATHS	MATHS		MATHS	MATHS
11.00	Business and O.R.	COMPUTING	Philosophy of science	Business and O.R.	COMPUTING
12.00	Presentation environment		Presentation of data analysis proj.	Statistics Theory	
2.00	SIMULATION THEORY AND PRACTICE	DATA ANALYSIS PROJECTS	GAMES AFTERNOON	ENVIRONMENT PROJECTS	Data analysis project
4.00	Photography	Graphic Design			COURSE CONFERENCE

THIRD YEAR

	Monday	Tuesday	Wednesday	Thursday	Friday
9.00	Control theory	Business and O.R.	INSTITUTION ATTACHMENTS	Control Theory	Attachment discussions
10.00	Statistics	Statistics		Statistics	Statistics
11.00	Economics	Current Affairs DEBATE		Project Discussions	Business and O.R.
12.00	MATHS	MATHS		Committees Brainstorming Interviewing	MATHS
2.00	PROJECTS	PROJECTS	GAMES AFTERNOON	PROJECTS	PROJECTS
4.00	EDITING	EDITING		EDITING	COURSE CONFERENCE

FIGURE 2: continued...

SECOND TERM

FIRST YEAR

9.00	Monday	Tuesday	Wednesday	Thursday	Friday
10.00	MATHS	MATHS	MATHS	MATHS	MATHS
11.00	COMPUTING	Electronics theory and design	Computing	COMPUTING	Instrument'n.
		Presentation Psychology	Presentation Electronics		Presentation Psychology
12.00	Institutions	Statistics	Statistics	Public speaking Interviewing	Statistics
2.00	Psychology Laboratory	Electronics Laboratory	GAMES AFTERNOON	Physiology Laboratory	STATISTICS LABORATORY
4.00	WRITING	WRITING		WRITING	COURSE CONFERENCE

SECOND YEAR

9.00	Surveys	Surveys	COMPUTING	Surveys	Surveys
10.00	MATHS	MATHS		MATHS	MATHS
11.00	Business and O.R.	COMPUTING	Business and O.R.	Business and O.R.	COMPUTING
12.00	Statistics		Presentation of data analysis project	Statistics	
2.00	SIMULATION THEORY and PRACTICE	DATA ANALYSIS PROJECTS	GAMES AFTERNOON	ENVIRONMENT PROJECTS	Data analysis project
4.00	Visual aids	Visual aids			COURSE CONFERENCE

THIRD YEAR

9.00	Control theory	Business and O.R.	INSTITUTION ATTACHMENTS	Control theory	Attachment discussions
10.00	Statistics	Statistics		Public Relations	Statistics
11.00	Economics	Current Affairs Debate		Project Discussions	Business and O.R.
12.00	MATHS	MATHS		Committees Brainstorming Interviewing	MATHS
2.00	PROJECTS	PROJECTS	GAMES AFTERNOON	PROJECTS	PROJECTS
4.00	EDITING	EDITING		EDITING	COURSE CONFERENCE

FIGURE 2: continued...

THIRD TERM

		Monday	Tuesday	Wednesday	Thursday	Friday
FIRST YEAR	9.00	MATHS	MATHS	MATHS	MATHS	MATHS
	10.00	COMPUTING	Theory & Art COMMUNICTN	Theory & Art COMMUNICTN.	COMPUTING	Instrumentat'n.
	11.00		Presentation industrial lab	Presentation industrial lab		Presentation Electronics
	12.00	Institutions	Statistics	Statistics	Public speaking Interviewing	Statistics
	2.00	Industrial Process Laboratory	Industrial Process Laboratory	GAMES AFTERNOON	Electronics & Instrumentation Laboratory	STATS LAB
						COURSE CONFERENCE
	4.00	PRINTING	PRINTING		PRINTING	
SECOND YEAR	9.00	State Comp Packages	State comp Packages	COMPUTING	State Comp Packages	State Comp Packages
	10.00	MATHS	MATHS		MATHS	MATHS
	11.00	Business and O.R.	COMPUTING	Business and O.R.	Business and O.R.	COMPUTING
	12.00	Statistics		Presentation of data analysis project	Statistics	
	2.00	SIMULATION THEORY and PRACTICE	DATA ANALYSIS PROJECTS	GAMES AFTERNOON	ENVIRONMENT PROJECTS	Data analysis project
						COURSE CONFERENCE
	4.00	Word Processing	Exhibitions Design			
THIRD YEAR	9.00	Accounting	Consulting	INSTITUTION ATTACHMENTS	Accounting	Attachment discussions
	10.00	Urban and Social studies	Writing papers, reports		Educational Systems	Government & Industrial Finance
	11.00	Economics	Business and O.R.		Business and O.R.	Communications Survey
	12.00	MATHS	MATHS		MATHS	MATHS
	2.00	PROJECTS	PROJECTS	GAMES AFTERNOON	PROJECTS	PROJECTS
						COURSE CONFERENCE
	4.00	EDITING	EDITING		EDITING	

TEACHING APPLIED STATISTICS FOR AND/OR IN INDUSTRY

H C Hamaker, Huygenslaan 11, Eindhoven

INTRODUCTION

The quality specifications a statistician who wishes to follow a career in industry has to satisfy have been detailed in a report prepared by a Committee especially established for that purpose by the American Statistical Association [1] and in a paper by A.A. Greenfield [2]. These qualifications are concerned partly with purely personal characteristics, which can not be taught, and partly with the statistical and technical knowledge needed or desirable. With these we will mainly be concerned.

To my mind the requirements listed raise the important question whether any person trained primarily as a physicist, chemist, or engineer with some secondary training in statistics is not better prepared for a career as statistician in industry than a person trained primarily as a statistician.

To assist in planning experiments in industry the statistician has (1) first to diagnose the problem which is often incompletely and/or incorrectly stated, (2) to take into account the technical possibilities and the circumstances under which the experiment will have to be carried out, (3) to issue clear and unambiguous instructions for performing the experiment, and (4) to present the results in a practical form by graphs or tables that can be read by non-statisticians, and are directly related to the problem the experiment meant to solve. In these activities technical insight and practical experience in experimentation are invaluable.

Further a physicist, chemist, or engineer with some basic training in mathematics can easily acquire the essential statistical knowledge from text books and the literature, if need be assisted by the extra-mural courses organised today by universities and statistical societies.

Last but not least, and perhaps most important, a person with practical experience in technology or the natural sciences will be better equipped to see through the idolatry that dominates the statistical science of the day. This will be my main topic.

At birth and in its early youth statistics was nothing but exploratory data analysis. Pearson's set of frequency curves were invented to describe observed populations which did not fit the normal distribution; student's t-test was derived to deal with small series of data in a brewery; and the design of experiments with the F-test had its roots in the complicated problems of agricultural experimentation. But since then the science of statistics has degenerated into a branch of mathematics, so much so that Exploratory Data Analysis, or EDA, is introduced as something new. Most of the literature nowadays consist in what Sir Claus Moser [3] in his Presidential Address to the Royal Statistical Society has termed "theories looking for data rather than real problems needing theoretical treatment". And the data used to illustrate such theories are then all too often wrested from their context and distorted in order to fit the theory, as some examples to be discussed further on will illustrate.

DESIGNING EXPERIMENTS I

One of the main functions of a statistician in industry is to assist in designing experiments; but sometimes no experiment is needed.

Example 1:

A factory developing a hot iron has trouble with the thermostat and asked for an experimental design. Five influencing factors were listed, but when I asked what was the influence of each of these, varied singly, my clients simply did not know. I therefore advised them to perform some one-factor experiments first; without some information it is not possible to design an experiment on a reasonable basis. My clients never came back because they had soon found out what was the source of their trouble.

Example 2:

An experiment was requested to find the sources of the variability in the "blowing time" of small electrical fuses, which was said to be far too large. The factors to be considered were: (1) the material, (2) the diameter, and (3) the length of fuse wire, (4) the current strength and (5) the type of current (direct or alternating), (6) the diameter and (7) the wall thickness of the glass tube, and (8) the nature of the gas filling (air or an inert gas). But when I asked "How large is the variability in the blowing time?" the answer was "Very large but we do not know more than that". So I wanted to see this first and instructed my clients to blow sets of fuses of one type at different currents and show me the results. The variability turned out to be quite reasonable except at one specific current, where there appeared a two-humped distribution. The physical explanation was evident; at this current the fuse wire was heated to near its melting point; for some that point was reached and these blew almost at once, while others stayed below the melting point and had to blow by evaporation and/or oxidation which is a much slower process. This simply was not known and threw quite a new light on the problem. No further experiments were needed.

Example 3:

A tool used to press small products out of a metal ribbon broke down too often and an experiment was requested to find the reason why? Factors were (1) the material of the tool, (2) the tool maker, (3) the hardening process, (4) the machine setter, and (5) the material on which the tool operated; but the difficulty was that these always blamed the defects on one of the others. I then advised to establish a group with representatives from all the departments concerned and jointly inspect a number of defective tools. It was then discoverd that 80% of the defects were due to one weak spot in the construction of the tools, which had not been realised before.

In such situations little or no statistics are required, only some common sense is needed. This I believe has to be taught by practical experience in the factories.

In teaching the design of experiments it should, in my opinion, be emphasized that the basic ideas originated in agriculture; where one has the spring and summer to carry out the experiment, and the autumn and winter to harvest and analyse the data, and plan another experiment for the following year. Working at that speed it is essential to draw the maximum of information from each experiment, and complicated designs with confounding, split plots, etc. are in order. But in industry the execution of a designed experiment will often not take more than a day, and seldom more than a week. And in that situation such complicated experiments including many factors are not the most effective solution. A series of experiments with only 3 or 4 factors may be much better. This has several advantages; one does not have to play all cards at once; one can gain experience in the subject matter as the series develops; and the results of a 3- or 4-factor experiment are much easier to explain than those of an experiment involving a larger number of factors.

Example 4

The following factors were to be included in a design to determine the visibility of light signals: (1) day or night, (2) the signal is continuous or flashing, (3) blinding of the observer by direct or indirect light, (4) colour: red, green, yellow, white, (5) distance: 100, 300 and 1,000 meters, (6) 11 observers. The responses were decided to be the intensity of the light source when the signal was judged: just visible, just right, and very clearly visible.

A full scale experiment with all factors crossed, yielding in all 3,168 data, had been carried out before I was consulted. Technically this may well have been the only possibility, because a special set-up had to be arranged, and the persons taking part could not be called off their work or off home too often.

An important question in this case would also be what the real purpose of the experiment is? If it is only to find conditions under which the light signals are invariably either easily or very clearly visible, then the data can be thrown into a computer and sifted out in some simple way. But if the purpose should be to express the responses as functions of the factor levels, and if the statistician had complete freedom to arrange the experiment as he likes best, then it might be much better to proceed carefully; for instance by first investigating the reproducibility of the responses taking into account weather conditions and fatigue as well; and then to study the relation between distance and visibility on a modest scale to find a suitable model to be used in further research; etc.

Example 5

A new type of radio tube developed in factory A gave a poor electron emission of the cathode when produced in factory B. In this case it is known at the outset that there exists a very large effect, the only purpose is to find its source, which is likely to be due to one small difference in production. A careful analysis showed that there were 18 steps in the production where the two factories slightly differed. experimentation is now very simple; the 18 conceivable factors can be split up in two blocks of 9. A set of tubes is produced with block 1 carried out in factory A and block 2 in factory B. This will at once reveal in which of these two blocks the source of trouble must be sought. That block can then be split up in the next experiment, etc. It turned out that for some reason the glass bulbs were in factory A rinsed out with water before going in production, while in factory B this was not considered necessary. The chemistry behind this phenomenon need not consider us here.

The conclusion to be drawn from these examples is that the actual planning of an experiment has many more aspects than are taught in text books on the subject. An experiment may have many different purposes, and may have to take account of technical restrictions which nearly always exist. Also the statistician has always to take a critical attitude with regard to the situation as it is represented by his clients, because these look at their problems from their preconceived points of view. For instance, if this is not mentioned by his clients, the statistician should, in the case of the light signals, bring up the problem of weather conditions, fatigue, color blindness, the use of spectacles, the selection of the observers, etc. In other cases he may have to reckon with the possibility of hidden or "lurking" factors, which have not been mentioned but appear to exist by a careful analysis of the data. B.L. Joiner [4] has collected a number of examples.

EXPERIMENTAL DESIGNS AND THEIR ANALYSIS

So much for my own experiences, let us now look at some examples taken from the literature.

Example 6 - taken from N.L. Johnson and F.C. Leone [5, p.135]
The response is the measured viscosity of a certain polymer, and the experiment includes four factors which are all nested within each other : U, V nested within U, W nested within UV, and Y nested within UVW; respectively with 2, 10, 2, and 2 levels. The analysis of the data is presented by the following analysis of variance :

TABLE 1

Analysis of variance of Example 6

As given by Johnson and Leone				Corrected	
Source	SS	DF	MS	DF	MS
U	0.07	1	0.07	1	0.07
V within U	503.48	9	55.94^{+++}	18	27.97
W within UV	454.72	20	22.74^{+++}	20	22.74^{+++}
Y within UVW	109.13	49	2.23	40	2.73

A check revealed that the SS's are correct but the DF's as assigned by the authors are wrong; the MS for V was overestimated by a factor 2 and turns out to be not significant.

But another feature is, that if W is nested within U the MS for U is suspiciously low; F(20.1) = 22.74/0.07 = 324 exceeds the 5% level of significance; this might indicate that W is not nested within U. This finally led us to carry out a complete analysis of variance with U, V, and W crossed, and leaving out Y as of minor importance. Result:

Analysis of variance considering U, V, and W crossed

Source	SS	DF	MS
U	0.07	1	0.07
V	488.00	9	54.22
W	12.09	1	12.09
UV	15.47	9	1.72
UW	1.08	1	1.08
VW	432.05	9	48.00
UVW	9.50	9	1.06

TABLE 2

The MS's for V and VW are of the same order which tallies with the view that W is nested within V, and that V itself does not contribute to the variance. But the MS's for U and all interactions within U are an order of magnitude smaller; and this can only be understood if U was crossed with both V and W.

The description of the experiment does not help: "... the original sample was subdivided, for ease of handling, into 2 random subsamples which can be represented by random variables U_1 and U_2. Each of these was then divided into 10 aliquots ($V_1 \dots V_{10}$). The first step in the analysis of the polymer was then carried out. Each aliquot was again divided into 2 (W_1 and W_2) and the second step of the analysis was performed." But what the steps in the analysis were and how they were carried out is not mentioned.

Anyhow something is definitely wrong; the experiment has been misunderstood, or the data have been misrepresented.

Example 7 - taken from C.R. Hicks [6, p.94]
The object of the experiment was to study the thrust force in drilling at different speeds, feeds, and in different materials. The data for the second material are reproduced below.

TABLE 3

Feed/Speed		100	220	475	715	870	The signs of $R_1 - R_2$				
0.004	R_1	192	136	122	108	136	+	+	+	+	+
	R_2	170	130	85	75	75					
0.008	R_1	386	333	318	472	499	+	+	-	+	+
	R_2	365	300	300	350	390					
0.014	R_1	810	779	810	893	1820	+	+	+	+	+
	R_2	725	670	750	890	890					

Now in presenting these data it is stated that "The order of the experiment was completely randomized", and this is at once belied by the signs of the differences between the replicates R_1, R_2. And what is meant by "completely randomized" taking into account that a drill gets blunted in the course of the operation?

Further it is stated that "the levels of all factors are fixed", which is evidently meant for the pleasure of the statistician, but for a technologist is perfectly meaningless. If further experiments of the same kind are contemplated the engineers will have complete freedom to choose different speeds, feeds, and materials.

Then one of the replicates at the end of the bottom row is, to say the least, suspect as an outlier, which occurs at the highest speed and feed; this might well be due to some instability, for instance when a drill gets blunted.

The data are given as an exercise, the 3 tasks imposed being:

1) Do a complete analysis and state your conclusions,
2) Plot any results which are significant,
3) Set up tests on means where suitable and draw conclusions.

From a technical point of view these questions do not make much sense; a complete analysis requires that we decide what to do with the suspected outlier first; a plot of the data is needed and is practically all that is needed, but plotting only significant results is impossible without plotting insignificant results as well. And the significant effects are visible enough and do not require a test.

Nevertheless such a set of data may be very useful in the classroom provided they are used to discuss what are the conceivable lurking factors that have not been mentioned, but should be taken into account or else left out wittingly. I have, for this specific case, listed a number of questions that might be asked elsewhere and I will not repeat them here: Hamaker [7].

Example 8 - again borrowed from N.L. Johnson and F.C. Leone [5, p.233] The response is the reflectance of a paint, the factors being different operations in the production of that paint, namely: (1) two liquids in which and (2) two different mills by which the pigment is dispersed, (3) three replicates nested within L and M, (4) two times of aging before the paint is finally mixed, (5) two ways of applying the paint, and (6) two wavelengths at which the response is determined; respectively L,M,R,T,S and W. T,S and W are crossed with L,M and R. The data are reproduced below.

Percentage reflectance of a paint:
L two Liquids, M two ways of Milling, T two aging Times,
S two ways of applying the paint, W two Wavelengths, R three Replicates

			L_1						L_2					
			M_1			M_2			M_1			M_2		
			R_1	R_2	R_3	R_1	R_2	R_3	R_1	R_2	R_3	R_1	R_2	R_3
T_1	S_1	W_1	44.5	44.5	45.0	47.0	50.5	48.0	44.0	44.5	44.0	47.5	48.5	46.5
		W_2	7.5	8.0	8.5	8.5	8.5	8.5	7.5	7.5	7.5	9.0	9.0	8.5
	S_2	W_1	43.5	42.5	44.0	45.5	44.5	45.0	43.0	44.0	43.0	44.5	43.0	43.0
		W_2	10.0	9.5	10.0	12.0	12.5	12.5	9.0	10.5	9.0	10.5	9.5	10.0
T_2	S_1	W_1	45.0	44.0	45.5	47.0	46.5	45.5	44.0	44.5	43.5	48.5	46.5	46.5
		W_2	7.5	7.5	8.0	8.0	8.0	8.0	7.5	7.5	7.0	9.5	8.5	8.0
	S_2	W_1	44.0	43.0	44.5	45.0	44.0	45.5	43.5	44.0	43.5	43.5	43.5	43.0
		W_2	10.5	10.0	11.0	12.0	11.0	11.0	9.0	9.0	9.0	9.5	10.5	10.0

TABLE 4

I first give my own analysis and then discuss that of the authors.

BEST COPY AVAILABLE

A) At a glance it is noticed that the level of the response is quite different for W_1 and W_2, and that the differences between replicates are considerably smaller for W_2 than for W_1. Hence it is best to treat the reflectances for the two wavelengths as two separate responses. Therefore the averages over the three replicates should be represented as in Table 5.

TABLE 5

Average percentage reflectances for W_1 and W_2

		W_1				W_2			
		L_1		L_2		L_1		L_2	
		M_1	M_2	M_1	M_2	M_1	M_2	M_1	M_2
T_1	S_1	44.7	48.5	44.2	47.5	8.0	8.5	7.5	8.8
	S_2	43.3	45.0	43.3	43.5	9.8	12.3	9.5	10.0
T_2	S_1	44.8	46.3	44.0	47.2	7.7	8.0	7.3	8.7
	S_2	43.8	44.8	43.7	43.3	10.5	11.3	9.0	10.0

Approximate 95% confidence intervals for these averages are:

± 1.0% ± 0.5%

B) The confidence intervals have been based on the pooled standard deviations derived from the sets of replicates, which are s = 0.86 and 0.44% for W_1 and W_2 respectively; this may not be quite correct, but serves to indicate the order of magnitude.

C) Some significant effects are obvious: the highest reflectances are obtained by the combination M_2S_1 for W_1, and by M_2S_2 for W_2. There may of course be other effects that are statistically significant, but it seems questionable whether these are worthwhile from a technical point of view, and the statistician in charge should first consult with the production manager or research department who requested this experiment; perhaps Table 5 is sufficient for their purposes.

D) Let us now look at the analysis by Johnson and Leone. They see this as an example of a split-plot experiment, the underlying model being stated as:

TABLE 6. The model for the data of Table 4

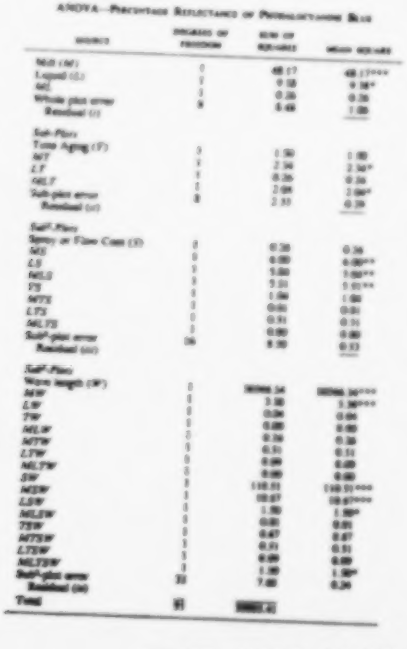
$$\begin{aligned} x_{mlqrst} = {} & A + M_m + L_l + (ML)_{ml} + \varepsilon_{q(ml)} && \text{Main Plots} \\ & + T_t + (MT)_{mt} + (LT)_{lt} + (MLT)_{mlt} + \varepsilon_{q(mlt)} && \text{Sub Plots} \\ & + S_s + (MS)_{ms} + (LS)_{ls} + (MLS)_{mls} + (TS)_{ts} \\ & + (MTS)_{mts} + (LTS)_{lts} + (MLTS)_{mlts} + (\varepsilon'_{q(mlt)} + \varepsilon_{(mlts)}) && \text{Sub}^2\text{-Plots} \\ & + W_w + (MW)_{mw} + (LW)_{lw} + (MLW)_{mlw}(TW)_{tw} \\ & + (MTW)_{mtw} + (LTW)_{ltw} + (MLTW)_{mltw} + (SW)_{sw} \\ & + (MSW)_{msw} + (LSW)_{lsw} + (MLSW)_{mlsw} + (TSW)_{tsw} \\ & + (MTSW)_{mtsw} + (LTSW)_{ltsw} + (MLTSW)_{mltsw} \\ & + (\varepsilon''_{q(ml)} + \varepsilon''_{q(mlt)} + \varepsilon''_{q(mlts)} + \varepsilon''_{q(mltsw)}) && \text{Sub}^3\text{-Plots} \end{aligned}$$

with a corresponding analysis of variance as shown in Table 7. The conclusion is drawn that there are strong interactions between L, M and S; that SW, MSW and MW are also significant; and that MT, MLT, MS, LS, MLS, LSW and MLTSW though significant are not very large. But no attempt whatsoever is made to explain what anyone of these interactions means from a technical point of view.

E) The interpretation as a split plot design is based on the assumption that the factors T, S and W not only correspond with a fixed effect but also introduce each a random error. The errors due to S are then nested within T, and those due to W nested within T and S; we should in that case expect the sub-plot error MS's to decrease from sub-plot (ii) to sub-plot (iv); as these are nearly equal the conclusion seems justified that T and S do not significantly contribute to the random error, and that the errors in the determination of the reflectance are the ones that matter.

TABLE 7

Analysis of variance of the data of Table 4 according to the model of Table 6, according to Johnson and Leone



BEST COPY AVAILABLE

F) But the split-plot analysis disregards the difference in the error variances of W_1 and W_2, which are mixed up. Moreover, if T and S contribute random errors it is conceivable that the error variances will differ between S_1 and S_2, and between T_1 and T_2. If these error variances are of practical importance they should be determined by experiments specially designed for that purpose, which can be of a very simple type.

G) Another problem is that the reflectance of 50.5% in the top row of Table 4 seems rather high in comparison with the values on both sides, so that it might be suspected as an outlier.

The mean range for the 16 sets of replicates for W_1 is $\bar{R}$ = 13.4%, and the range for the top group under L_1M_2, which is R = 3.5, falls just outside the 3σ limit, which confirms our suspicions.

Further, for the combination $L_1M_2W_1$ we have the following sets of replicates:

Analysis of variance							After replacing 50.5 by 47.0	
					MS	DF	MS	DF
T_1S_1	47.0	50.5	48.0					
T_1S_2	45.5	44.5	45.0	Treatments	8.61	3	4.18	3
T_2S_1	47.0	46.5	45.5	Replicates	0.15	2	0.44	2
T_2S_2	45.0	46.5	44.0	Residuals	1.51	6	0.44	5

If we reject the value 50.5% as suspicious its replacement values is found to be 47.0%; and, as shown, the replacement has a considerable effect on the MS's in the analysis of variance. Following this trail one step further it is found that by this replacement the residual error variances in the analysis of variance of Table 6

which are	1.06	0.29	0.53	and	0.24
change into	1.11	0.13	0.40	and	0.15

respectively. This will not materially change the conclusions with regard to the significant effects in Table 6, but it illustrates the influence a single suspected datum can have on the error estimates.

H) I also believe that an experiment as that of Table 4 should not in an industrial environment be called a split-plot experiment. Split-plots belong to agriculture where they came from; but for industrial purposes the design illustrated by Table 4 is more correctly described as a full 2^5 design with factors LMTSW and with 3 replicates nested within L and M but crossed with T, S and W; and it is only a split-plot design when T, S and W not only represent fixed effects, but are also considered as sources of random error components.

A STATISTICAL IDIOSYNCRASY

To me the statistical analysis presented in Table 6 illustrates an idiosyncrasy prevalent among statisticians, the chief symptom of which is the canonization of tests of significance. No attempt whatsoever is made to present the results of the experiment in a practical form; the only purpose of the experiment seems to be to test significances, and thereby the problem is considered solved.

One reason for believing that a training as physicist, chemist or engineer might be preferable as a basis for a statistical career in indstry is that this training may render people less susceptible to that idiosyncrasy. And one of the purposes of preparing statisticians for careers in industry should be to inoculate them; or to disenchant them if they are already infected. Here is a list of views they should be made familiar with.

a) Standard deviations σ are never known, except perhaps with simulated data. all we can dispose of is some estimate $\hat{\sigma}$ which we are prepared to use as if σ were known. But even if that estimate is based on 50 degrees of freedom, which is quite a lot, it may still be 10%, and occasionally 20%, in error; and under repeated use of that estimate the level of significance of a test may lie anywhere between 90% and 98%, when we pretend it to be 95%.

b) And this defect cannot be remedied by having recourse to the t-distribution; for that distribution presupposes that every t is based on a fresh estimate of σ; if we use an estimate s repeatedly in a series of t-tests the level of significance may be as much in error.

Hence with a table for the t-test it should be explained that the underlying assumption is, that the estimate s in t is used only once; and that, if the same estimate is used repeatedly the level of significance is only approximate. Otherwise we are cheating.

Of course the experience shows that this does not matter in the least in statistical practice; the real value of a significance test does not lie in the exact value of the associated probability; it resides in the fact that such tests provide us with simple and universally accepted rules for reaching decisions.

c) The normal distribution should be taught in the same vein, as an accepted convention for arriving at conclusions we are willing to abide by as long as this convention is unchallenged.

d) In the analysis of an experiment every mean square corresponds to an effect, and these effects, and not the mean squares, should be earmarked according to their degree of significance. For example in the case of Table 7 the effects are

Grand Mean ; 27.03, W : -17.84, M ; +0.71, MW : -0.19

and the average reflectances for the combination of M and W are

$$M_1W_1: \ 27.03 + 17.84 - 0.71 - 0.19 = 43.97\% \ ,$$
$$M_2W_1: \ 27.03 + 17.84 + 0.71 + 0.19 = 45.77\% \ ,$$
$$M_1W_2: \ 27.03 - 17.84 - 0.71 + 0.19 = \ 8.66\% \ ,$$
$$M_2W_2: \ 27.03 - 17.84 + 0.71 - 0.19 = \ 9.71\% \ ,$$

as it can be verified from Table 4.

The mean squares for W and for MW, respectively 3056.63 and 3.38, are both marked as highly significant, but it is very questionable whether the effect of MW = 0.19% has any significance at all from an industrial standpoint; and an analysis of variance giving only the mean squares without stating the magnitude of the effects leaves out the information that matters in the first place.

Also the effect of MW is only about 1% of that of W, and when such drastic differences occur it may always be simpler and more effective to split up the analysis in parts corresponding to the levels of the factor with the largest effect; as we have done in Table 5.

e) The actual planning of an experiment can best be learned in the factory. The situations in a chemical industry, a steel industry, a textile industry, an electronic industry, etc. are all different. If there is a statistical department in the industry new members of the staff can be taken along and learn from their experienced seniors. In Dutch universities it is also a common practice to place students for a training period of about half a year in some industry before their final exam, and this is very helpful. I always made it a point to visit the factory departments that requested a designed experiment and take one of my assistants along with me; you have to see what is going on, for example in order to become aware of possible complications and lurking factors. Effective teaching of the planning of experiments in the classroom is to my mind much more difficult; perhaps a classroom discussion of such a situation as that of Example 4 (visibility of light signals) and all the many possible ways to organise a restricted experiment for a start might be useful.

f) In industrial experiments outliers do occur and have to be reckoned with!! In a recent document on procedures to validate the performance of methods of analysis B. Horwitz [6] states that collaborative test programs (in which a number of laboratories take part), "seem to generate an inherent 5 to 15% outliers; above 20% without an explanation (e.g. failure to follow the method) is ordinarily excessive. The sooner an outlier is investigated, the more likelihood of finding an explanation". In keeping with this statement chemists often carry out their analyses in duplicate, and the same practice may be recommended for industrial experimentation, at least if the tests are not too expensive. In the case of Example 8 this could have eliminated the uncertainty concerning the one suspect reflectance of 50.5%. With duplicate tests we should then start by investigating the differences between duplicates before embarking on a full scale analysis of variance; or the ranges or standard deviations of series of more than 2 tests. Also it would be advisable as a preliminary study to determine the repeatability of the test method, in order that during the execution of the final experiment a criterion is available by which the occurrences of outliers can be judged while the tests are made.

Significance tests for outliers are of little use. Barnett and Lewis [9] list no less than 43 different tests for a single series: σ and μ both known, one of them unknown, both unknown, one upper outlier, two upper outliers, one or two extreme outliers, etc. This is all due to the exalted importance attached to the 5% and 1% level of significance. Such a variety of tests is, however, only confusing.

From another chapter in Barnett and Lewis it is seen that detecting and testing outliers in the data from designed experiments is considerably more difficult. The case is further complicated because in an experiment of some size there may easily be more than one outlier and we do not know how many.

Nevertheless, outliers must be reckoned with and it is surprising to find that text books on the design of experiments pay no attention to this problem; in many of them the term "outlier" does not even occur in the index.

We can distinguish three types of outliers: (1) those that can be technically explained, (2) those that are so outrageous that there can be no doubt, and (3) the doubtful cases for which there is no technical explanation; these last are the real source of trouble.

The detection of suspect data will in the first place depend on a critical examination of the data or of residuals. In doubtful cases there is in my mind only one reasonable way to deal with outliers. After concluding his analysis the statistician has to draw conclusions and make recommendations. In the face of possible outliers the analysis should then be carried out with the suspected data in place and after rejecting and replacing them. If the conclusions and recommendations are unaffected there is no problem, but if they are materially changed this should be stated so that it can be taken into account.

Anyhow "detecting and dealing with outliers" should form part of the training of statisticians who aspire after a career in industry.

6) On the other hand they should be warned against Bayesian theories. Every successful industrial statistician has to be a Bayesian in the sense that he has (1) to make effective use of past experiences, (2) to consider the costs

of further research, and (3) to take into account the economic consequences of his recommendations. Conceptually there may exist some economic optimum. But as soon as we set ourselves the task of reaching that optimum, we first have to define that optimum, and this invariably leads to unrealistic assumptions and illusive solutions.

In sampling inspection for instance, Bayesian theories have to force past experience into a stationary a-priori distribution, and as costs they only consider the cost of inspecting one item and the losses due to accepted defectives. This is all wrong, because the purpose of sampling inspection is not to prevent the acceptance of defectives but to prevent the production of defectives. A producer who knows that his lots will be inspected will pay more attention to the quality of his products; it is the interaction between sampling inspection and product quality that is important. Bayesian theories may easily lead to solution under which 50% or so of the lots received have to be rejected; a condition utterly impossible in industry, because it will upset the regular production process and stock policy.

Conceptually the problems a statistician has to solve, or assist in solving, may have an economic optimum solution, but this solution must be accepted as lying out of reach. By his work a statistician can help to achieve improvements and with that he has to be satisfied.

h) Last but not least, statisticians should stop fooling themselves and cheating others; and should learn to see applied statistics in its true light.

By producing a standard for sampling inspection by variables based on the assumption that σ is known we are cheating, if we do not specify what we mean by that.

By defining the AQL as "satisfactory as a process average" we are cheating; the probability of acceptance at AQL is only 90% for some of the sampling plans in the Mil.Std.105D, and in that case a producer has to establish a process average considerably better than the AQL to stay out of trouble.

To wait until 2 out of 5 successive lots have been rejected before switching to Tightened Inspection is absurd, at least at the present time. Every rejected lot is a source of trouble and will stimulate some action; and the quality requirements of electronic components are so high, that even a single defective may bring in its train an inquiry as to the source of the defect and the steps taken to prevent it from recurring.

In two standards, ISO 5725 and ISO 4259, the repeatability r and the reproducibility R of a standard method of test are stated to be associated with a probability level of 95%; but as they are only estimates based on a limited number of degrees of freedom and on a set of assumptions which we know very well can never be quite true, it would be much more honest to state that the associated probability is "of the order of 95%".

In ISO 4259 a transformation is applied to the data which is quite superfluous, and a relation between a set of standard deviations s_i and corresponding means $\bar{x}_i$, which by a least squares fit should be $s = a\,(\bar{x})^{0.574}$ is arbitrarily changed into $s = a(\bar{x})^{2/3}$, whereby a systematic bias of the order of 60% is introduced in the resulting repeatabilities r. As a statistician I protest, but as a practitioner I am inclined to recognize that, despite these statistical imperfections, the standard satisfies practical requirements. The field of application of the repeatability and reproducibility is too insensitive to bother about such statistical refinements.

Perhaps my views are best expressed by saying that statisticians are too finicky, fastidious, or pedantic about their levels of significance; and this may not only hold for industry but equally well elsewhere.

A MOTTO TO CONCLUDE THIS SERMON

After reviewing the statistical theories of his day Walter Shewhart [10] introduced his famous 3σ limits for control charts in the following words:

"Experience Indicates that $t = 3$ Seems To Be An Acceptable Economic Value"

There is no mention of a normal distribution !

There is no level of significance attached !

It is recognized that we are at bottom facing an economic problem !

For which we have to be satisfied with an acceptable solution !

Based primarily on experience !

It is in this spirit statistical practitioners for industry should be trained !

References

[1] Preparing Statisticians for Careers in Industry: Report of the ASA Section on Statistical Education Committee on Training Statisticians for Industry (1980) The American Statistician, Vol.34, No.2.

[2] Greenfield, A.A. (1979) Statisticians in Industrial Research: The Role and Training of an Industrial Consultant. The Statistician, Vol.28, No.2.

[3] Moser, Sir Claus (1980) Statistics and Public Policy. Journal of the Royal Statistical Society A, 143, 1-33.

[4] Joiner, B.L. (1981) Lurking Factors. The American Statistician, Vol.35, No.4.

[5] Johnson, N.L. and Leone, F.C. (1964) Statistics and Experimental Design in Engineering and the Physical Sciences, Vol.II, Wiley, New York.

[6] Hicks, C.R. (1964) Fundamental Concepts in the Design of Experiments. Holt, Rinehart and Winston, New York.

[7] Hamaker, H.C. (1971) New Techniques of Statistical Teaching, Review of the International Statistical Institute, Vol.39, No.3.

[8] Horwitz, W. (1982) Outline of Collaborative Study Procedure to Validate Performance of a Method of Analysis. Prepared for the Committee on Collaborative Studies of the Association of Official Analytical chemists.

[9] Barnett, V. and Lewis, T. (1977) Outliers in Statistical Data, Wiley, New York.

[10] Shewhart, W. (1931) the Economic Control of Quality of Manufactured Product. Von Nostrand, Princeton.

PRACTICALS, CASE STUDIES, PROJECTS AND PROFESSIONAL EXPERIENCE: SOME METHODS OF DEVELOPING SKILLS AMONG BUDDING STATISTICIANS

G K Kanji, Sheffield City Polytechnic

INTRODUCTION

Over the last 50 years, in response to the increasing demand for statisticians in research, industry and commerce, the teaching of statistics in the British Academic institutions has expanded greatly. Most Universities, Polytechnics and Colleges have either a department of statistics or a statistics section to provide various courses in statistics.

The demand for statistical education at the academic institution comes broadly from three classes of students:-

a) those aiming at becoming academic statisticians i.e. future teachers of statistics,
b) those aiming to be applied (practising) statisticians (i.e. industrial statisticians, social statisticians, etc.)
c) those merely requiring numerate training.

In addition to the above requirements there is an increasing amount of statistical education in schools.

The theme of this session is "Training of Industrial Practitioners" (i.e. training budding statisticians to be industrial practitioners) and I will therefore try to explore some of the methods available to academic statisticians to help them train as well as teach the future applied statisticians. From the outset this teaching and training of future applied statisticians requires two things. Firstly the provision of an appropriate core of theoretical knowledge and secondly assistance with the development of appropriate skills in order to apply this knowledge fruitfully. In short we have to teach "Applied Statistics". Many people will be asking whether it is possible to train applied statisticians at an academic institution and if so, how? My reaction to these questions will be that it is not possible in the real sense but we may go some way towards this goal. Let us therefore explore how far down the road of the development of skills we may go and illustrate the methods available to academic institutions.

By tradition, academic statisticians over a period of 50 years have developed academic works with great expertise and excellence. It is therefore no problem whatsoever to provide budding statisticians with theoretical knowledge (i.e. to teach) but we encounter difficulties when we try to help them to develop their practical skills (i.e. to train them) to solve the types of problems which exist in various (applied) environments. Part of the difficulty lies in the fact that academic statisticians throughout the world do their jobs so well that they often find it almost impossible to change. It is ironic that if what they do well is no longer wholly appropriate to the world of which they are part then their very proficiency makes it more difficult for them to change. However, in spite of my unbounded admiration for the academics, I feel that some of us (academics) are not completely fulfilling our roles as educators in an applied subject.

Yates and Healy (1964) discussing about the contents of courses for professional statisticians suggested that the course should be judged on the following criteria:

(i) how far the techniques to which it gives rise are likely to be required by practising statisticians

(ii) how far its mastery contributes to general understanding of statistical reasoning and improvement of statistical skill

(iii) the ease with which it can be mastered.

Speaking about the training of statisticians for economics and business Dean (1964) said that "the students should in their course, be taught not only how to formulate a problem or construct an econometric model, not only how to draw sound conclusions for themselves, but also to assess the findings in everyday language".

Presenting his views about the training of government statisticians Bishop (1964) said that "the relationship between theory and practice, and a feeling for the logic of this bridging process and the usefulness of models, can only be acquired by using crude, real life figures - preferably a lot of them - and hammering out some conclusions, under guidance of course. Ideally syndicate or project work would provide opportunities for team work, discussion of principles, writing up results and perhaps for computer work".

Jeffers (1976) talking about teaching said that "the formal teaching will never be enough to provide the necessary knowledge and experience for an applied statistician. Our teaching must be allied to a form of practical training which builds on, and widens the experience gained through formal education".

Spicer (1964) expressing his doubts about the training of medical statisticians wondered "whether the teaching of statistics is not more heavily weighted by considerations of rigour and mathematical technique than, say, engineering".

At this point it is worth reiterating the suggestions on the content of courses made by the I.S.I. report on the integration of statistics (1980), namely

a) better balance between theory and practice,

b) the illustration and exercise from real life situations in the classroom setting,

c) assistance from outside in the development of case materials which can be introduced into the curriculum,

d) joint programmes to establish the relationship between theoretical training and applications in order to yield case studies for course work, and

e) the development of the special purpose teaching materials to bridge the gap between academic training and applied statistical work.

In another report, prepared by the A.S.A. section on Statistical Education Committee on training of statisticians for industry (1980), recommended

i) inclusion of a wide variety of statistical techniques in the course work,

ii) the provision of experience of real problems through a co-operative programme, and

iii) the development of communciation skills as part of the graduate curriculum.

In describing the philosophy of an applied statistics course Kanji and Gilchrist (1981) pointed out that "the aim of the course is to provide an education for the future applied statistician as distinct from an education in statistics. Thus the aim is not just to impart a knowledge of statistics to a student but to impart the necessary skills that are required to make use of that knowledge as an effective applied statistician".

It is therefore very clear from these views, reports and Kanji and Gilchrist's idea of a degree in applied statistics that skills in application are crucial in the role of the statistician and that he is useless to any employer without some problem-solving ability. Hence we will have failed as educators if we ask the employer to provide all the opportunities for problem-solving skills. In the following sections I would like to illustrate and discuss some of the methods that are possible to practice within the academic framework and which I believe help trainee statisticians to develop the skills they require in order to do their job efficiently, namely practical work; case studies; projects and professional training.

PRACTICAL WORK

We all know that statistics is an applied subject which lives on the information collected from other disciplines such as Biology, Agriculture, Medicine, Industry, Public Health etc. For the analysis and interpretation of such information (data) it is necessary to call for assistance from the theory of statistics which is in turn based on the science of mathematics together with the art of good reasoning. This gives the subject a hazy nature which the student, in the early stage of his statistical studies, finds rather confusing. However, the student new to statistics may be helped to iron out his confusion by conducting a series of simple experiments. The collection of information for and the analysis of data obtained from these experiments offers an experience of some of the difficulties encountered in more sophisticated statistical analyses. By meeting and attempting to overcome these difficulties, the student gradually builds up his own confidence and ability and it is for this reason that the need for statistical practical work is of prime importance.

In describing the role of a statistical laboratory in the teaching of statistics Kanji (1974) has mentioned that "the greatest difficulty for students in the learning of statistics is the transfer of thoughts i.e. from their basic knowledge of mathematics to basic ideas of statistics". Statistical practicals help this process and are necessary for developing a better understanding of the theoretical aspects of the subject. Such practicals should not of course be seen as isolated events but rather as part of a sequence of practicals gradually increasing in sophistication. Therefore the provision of facilities for conducting experiments, the purpose of which is the collection of data and their analysis, is an important role of statistical education.

There are at least two ways one can present practical problems in the classroom. Firstly by asking students to conduct simple experiments and collect the data and then analyse it. A typical such example, taken from our Statistical Laboratory Manual, is given below.

STATISTICS LABORATORY Experiment No.32
Statistical Concept - Linear Regression
Approximate time: 30 minutes No. of students: 2

AIM The experiment is designed to illustrate the method of fitting a regression line to data generated by the student.

EQUIPMENT One toy 'hotwheel' type car, a sloping length of track and a tape measure.

EXPERIMENT
(a) Hold the car steady on the line drawn 5cm from the bottom of the track - then let it go. Measure carefully the distance it travels from the end of the track to where it comes to rest.
(b) Repeat this with the car starting on the 10cm, 15cm, 20cm, 30cm, 40cm, 50cm and 70cm lines in turn, making eight readings altogether.

ANALYSIS
1. Summarise your results in the form:

Distance up the Track (x)	5	10	15	20	30	40	50	70
Distance Travelled (y)								

2. Draw a scatter diagram of the data.
3. If 'n' is the number of readings, we have
$n = 8,\ \Sigma x = 240,\ \Sigma x^2 = 10650$

4. If x and y are connected by a relationship of the form $y = a + bx$ (the regression line) then: $\Sigma y = na + b\Sigma x$ and $\Sigma xy = a\Sigma x + b\Sigma x^2$ Substitute the values obtained in (3) into these equations and hence solve them for 'a' and 'b'. State the equation of the regression line and superimpose the line on your scatter diagram in (2).

5. Using your regression line, estimate the distance the car would travel from (a) the 60cm line and (b) the 90cm line and then try it out in practice, and compare the agreement between the predicted results and the experimental ones.

State clearly your conclusions as a result of the experiment, commenting particularly on whether or not you feel an assumption of a linear relation was correct.

Secondly, a variety of practical problems which have been collected from industrial organisations, government establishments and research centres could be considered. Some will be relatively short and will be worked by students on an individual basis, others may take several hours or even last for more than a week and will be tackled by groups of students. In some cases the problem may involve only one technique but most will bring together ideas from different parts of a theory course. A typical practical problem of this kind is as follows.

PRACTICAL

Title: Some Problems Concerning Mould Temperatures

An investigation was carried out on a five section double gobbed IS machine to study the relationship between the operating temperatures of the mould and dimensional quality of the bottles produced.

It was anticipated that the rate at which glass is cooled by different parts of a mould during forming would have some effect on the shape and glass distribution in the bottles produced.

The mould surface temperature is known to be one process variable which affects the rate of heat transfer from the glass and has been shown by previous measurements to vary considerably between the moulds on an IS machine.

A hand fibre optic pyrometer was used to measure the mould surface temperatures.

Fig.1 illustrates the positions on the blank moulds at which the temperatures were measured. Readings were taken at each of these positions for both the inner and the outer mould. The mould surface temperature measurements are given in Table 1. The wall thickness of the bottles, measured using callipers after annealing, are given in Table 2. This was made on the flat sides of the bottles (Top, Middle and Bottom), as illustrated in Fig.2.

Analyse the data in detail and write a report discussing whether mould temperature may have a useful relationship with the occurrence of rejectable ware.

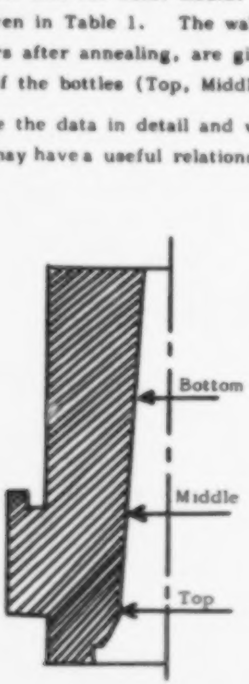


FIG.1 Positions on the Blank Mould Halves at which the temperature measurements were taken.

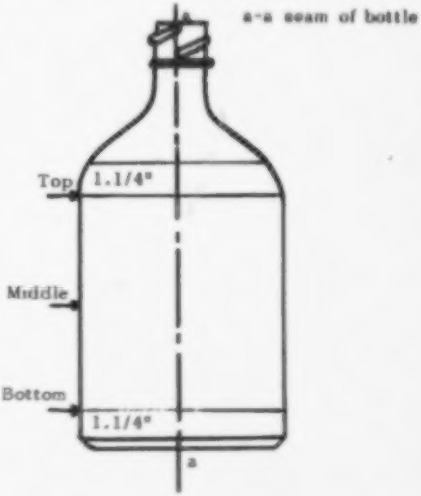


FIG.2 Positions on the bottle at which wall thickness measurements were taken.

TABLE 1

Blank Mould Surface Temperatures in ^{o}c.

		Section 1		Section 2		Section 3		Section 4		Section 5	
		LHS	RHS	LHS	RHS	LHS	RHS	LHS	RHS	LHS	RHS
Inner	T	473	484	442	491	426	392	440	442	409	418
	M	521	506	455	497	430	385	491	478	466	430
	B	485	479	449	461	410	382	466	460	452	424
Outer	T	518	517	474	462	417	430	404	409	402	414
	M	547	547	491	489	446	434	449	464	443	449
	B	527	528	468	464	434	410	454	463	436	426

TABLE 2

Wall Thickness Measurements of the Bottles (in inches)

		Section 1		Section 2		Section 3		Section 4		Section 5	
		LHS	RHS	LHS	RHS	LHS	RHS	LHS	RHS	LHS	RHS
Inner	T	.131	.080	.125	.120	.105	.120	.128	.120	.095	.110
	M	.106	.070	.112	.102	.088	.104	.105	.095	.095	.096
	B	.130	.080	.120	.108	.088	.130	.130	.106	.111	.112
Outer	T	.110	.100	.100	.110	.090	.125	.095	.122	.095	.120
	M	.098	.100	.080	.120	.082	.130	.086	.124	.095	.090
	B	.085	.098	.095	.140	.090	.125	.092	.110	.092	.120

CASE STUDIES

The importance of case studies in the study of applied statistics is that they help the student to understand the uses and abuses of statistical techniques and their applications. The most obvious benefit is that they can be used to demonstrate the application of statistical reasoning and techniques in an applied area without having to wait for a long period of time encountered by a real life problem. A further point is that whereas periods of professional experience (which I will discuss later) are restricted to one or two firms and hence in general to a limited number of problems, case studies may be drawn from a wide variety of areas. Thus the future statisticians can be made aware of problems in a variety of different areas including social, business, government, marketing, biometrics and so on.

In describing the use of case studies in teaching statistics Bissell (1975) has pointed out that case studies can be presented in different ways. "At one extreme, the problem and its solution may simply be described by the expert responsible for dealing with it. Whilst this may provide a fascinating (if not bewildering) glimpse of the artist at work, it often fails to involve the audience in the situation. At the other extreme, the problem may be formulated and then tackled by syndicates representing board meetings or management committees. The main difficulty here is that the leader or presenter of the study has to steer the discussion towards a solution that he knows to exist; rather than evaluate possible alternative actions that might have been either disastrous or even more successful than that actually adopted".

In spite of the difficulties I believe that the presentation of case studies similar to the latter method mentioned by Bissell is much more useful to students because it gives them an opportunity to participate and work as a group to solve the problem. It may be worth remarking that this method is the harder of the two for the presenter i.e. for the academic statistician. A typical case study which could be used by the students either singularly or in a group is as follows.

CASE STUDY

Title : Replacement of a Wearing Part

a) A part on a machine is subject to wear and has to be replaced regularly. As it ages the part tends to break more and more of the items produced by the machine. Thus the part incurs two types of cost - cost of replacement of the part, and cost of breakages. It is required to minimise total cost. Explain how such minimisation can be achieved for each of the following situations.

i) Where the number of breakages depends directly upon the age (in days) of the part.

ii) As (i), but with the condition that the supplier of the part, by way of guarantee, agrees to cover the cost of all items broken in the first 'k' days of usage of the part.

iii) Where the number of breakages appears to be a quadratic function of age (in days) of the part.

In any explanation, denote the cost of replacement of the part by C_R(£), the cost per breakage by C_B(£) and the number of breakages during the t^{th} day from time of installation of the part by n_t.

b) The part to be replaced at regular intervals is available from two suppliers - A, who charges £55 per part for his brand, and B who charges £50 per part for his particular brand. The cost per breakage is 50p. As the result of controlled trials the firm producing the items has available the following data on breakages.

Day from time of installation of part	Breakages during that day Brand A part	Breakages during that day Brand B part
5	2	2
10	4	5
15	4	7
20	7	8
25	8	10
30	11	14

Recommend the brand the firm should buy, and how often the part should be replaced in this case.

Discuss the possible effects of randomness in the data on the recommendations.

Suppose the supplier whose brand has not been recommended is willing to offer a guarantee to cover the cost of breakages during the first 5 days of usage of the part. Examine the effect of this guarantee, if any, upon the recommendations previously made.

Note : The total number of items produced each day can be taken as fixed (constant).

c) Suppose the part is available from a third supplier - C, who charges only 140 per part for his brand. Whilst brand C is known to perform well during the early days of its usage (causing few breakages), it is also known to produce more breakages as it ages than either brand A or brand B. The following data concerning breakages is available for analysis.

Day from time of installation of part	Breakages during that day for brand C part
5	1
10	5
15	8
20	14
25	26
30	34

Consider the effects of this information upon any recommendations previously made.

Just as the practical programme allows the student to progress from relatively simple to relatively sophisticated materials so will the case study programme.

PROJECTS

The inclusion of statistical projects in any training programme for statisticians is vital. Many academic institutions are not very keen to introduce project work at the undergraduate level as part of their curriculum, reasons for this unwillingness being the difficulties in generating enough appropriate and distinct projects and the necessary supervision.

In describing the aims of the project Kanji (1979) said that "the project is the most personalised aspect of the course in statistics and aims to provide students with the opportunity to apply and integrate the knowledge gained during the course by an extended study and to develop an organised and methodical approach. It also provides them with the opportunity

(a) to improve their ability to write critically and exercise judgement in a chosen area of statistics,

(b) to exercise their skills in reviewing and condensing published information,

(c) to develop their powers of written expression,

(d) to assist them in the use of the library and the various sources of literature,

(e) to help maintain a continuous thread in the main branches of statistics".

A major aspect of project work is that it makes it necessary for the student to go beyond the scope of the work covered during the more conventional aspects of their course work.

Clearly these aims are interlinked and it is therefore expected that an undergraduate student might take the following steps

(i) understand and define the problem clearly

(ii) organise the work in an effective manner

(iii) extend his/her knowledge through guided reading if appropriate to the development of a solution

(iv) interpret results and draw valid conclusions, and

(v) present his/her work in the form of a logical and coherent report.

It is difficult to put forward a strict rule about the suitability of a project because projects of a fairly broad nature could meet the above aims just as adequately as highly specialised ones. However, emphasis should be put on the maintenance of continuity in one of the branches of statistics and on a thorough understanding of the statistical analysis involved in the project.

A typical example of a project proposal for an undergraduate student might be the following.

PROJECT PROPOSAL

Title of Project : Environmental Impact Analysis

Aims of Project :

A survey was carried out by the Greater Manchester Passenger Transport Executive into the attitudes of households in the Hilcock area concerning the environmental impacts caused by bus services. The residents were asked to indicate the possible environmental impacts which bothered them. The results obtained indicate that though traffic in general is a greater source of bother than bus services in particular, the following aspects of bus services are a source of bother to at least 20% of the residents in the sample in Hilcock.

a) Risk of accident

b) Noise

c) Vibration

d) Dust and dirt

The purpose of this work is to set up a model to investigate the factors (i.e. characteristics of householder) affecting the level of concern within each of the various aspects of bus services.

Sequence of project

1. Survey of relevant literature
2. Familiarisation with the detailed theory of the Log-Linear models and possible modelling procedures
3. Familiarisation with GLIM
4. Fitting a Log-Linear model using GLIM
5. Analysis
6. Conclusions and Report

Resources available - Data and GLIM Manual
Supervisor - G K Kanji

PROFESSIONAL EXPERIENCE

An applied statistics course should be designed to develop the students' understanding of statistics and of associated methods and their ability to apply this knowledge to problems related to the application of statistics in industry in commerce and in government etc. One way to achieve this is to develop a course which includes a period of professional experience.

In describing aims and objectives of professional experience Jones and Kanji (1980) have pointed out that the main aim is to use work experience so that students can set their theoretical knowledge into perspective. The objective is to allow the student to work for an organisation within which he or she is performing a useful and necessary function in a problem solving capacity, thus developing his working skills both technical and personal.

Oakes (1981) has echoed Jones and Kanji's sentiment by describing the aims and objectives of professional training as follows.

1) To integrate theory and practice and to reinforce learning.
2) To enable students to relate their specialism to the wider social context and to bring their critical faculties in a non academic environment.
3) To increase knowledge of subjects outside of the academic course and of the world outside the University.
4) To provide a supplement to or a substitute for, some University practical work and to enable the student to learn and to appraise practical skills.
5) To gain insight into personal and group relationships in the employment situation.
6) To increase the maturity of the student.

7) To clarify and enhance career prospects.

8) To improve the students' ability to communicate orally and in writing.

9) To provide opportunities for the exercise of judgement and improve powers of decision making. Also to be able to meet deadlines.

The professional experience may be regarded broadly as an extension of the continuous assessment of the student into the industrial period. The student is therefore required to write a report at the end of the period and this is read and commented upon by the tutor who made the visit. The purpose of the report is two-fold, one to gain experience of report writing and two to illustrate to the student that all aspects of working life have to be accounted for by some means.

In order to give you some idea about a typical students' activity while he/she is having a professional training I would like to provide you with a very brief summary of a full student's report.

SUMMARY OF INDUSTRIAL PLACEMENT AT CIBA-GEIGY, HORSHAM, JANUARY-SEPTEMBER 1981

This particular student placement was with Ciba-Geigy. It is one of the largest multi-national companies in the world which originates, manufactures and markets organic chemical products. The Group is made up of affiliates in some fifty countries, the parent company being based in Basle, Switzerland.

The student was involved with the pharmaceutical area of the company at Horsham working with a group of ten statisticians.

The work done in the statistics department is mainly concerned with clinical trials on drugs, in phase three and four, that is licence for studies to be carried out on human beings. The trials are designed by the department in consultation with the medical department and the record forms are also designed by statistics. This ensures that the trial can answer the initial questions set out by the medical department with the appropriate statistics. The design of the trials are usually double-blind with a within patient crossover including a placebo for control. This gives rise to some interesting analysis including analysis of variance, transformations, t-tests, Hotelling T^2, basic statistics such as frequency counts, means, modes, median standard error; curve fitting etc. with which he was involved on most projects.

An interesting part of the students placement was a visit by Professor V. Barnett from the University of Sheffield. He gave a seminar on Outliers in Statistical Data. This showed how the department was not only concerned with pumping out figures, but interested in continuing updating their knowledge.

The computer facilities were very good as the department has linked to the company IBM 4341 computer which gave them the ability to use many different languages or statistical packages such as SAS (statistical analysis system), STATPK, GLIM, BMD through a Librarian system. The statistics department also has two microprocessors with plotter and printer linked to them and also a word processor for the design of record forms.

The two placements with this firm have given the student an opportunity to widen his theoretical knowledge. It has also given him the experience of data handling and analysis using computers. Finally he has gained some administrative experience in preparing data and co-ordinating data preparation.

It is clear that the role of the industrial supervisor, who is after all an experienced industrial practitioner, is crucial.

CONCLUSION

The practical work, case studies, projects and professional experience therefore should show the students how statistical ideas, mathematical techniques, computer usage and communication together with common sense, are all necessary to find a reasonable solution to practical problems. The object is to show that there is often no 'right' answer but that the applied statistician needs to be able to find the best possible method with the given resources bearing in mind the existing constraints. The message that there is often no 'right' answer should become apparent to the students during the continuing case study programme and should be well established by the time they undertake their project.

The author believes that under the present educational system inclusion of practical work, case studies, projects and professional experience in statistics courses at the undergraduate level can bridge the gap between theoretical statistics and real world statistical problems. This in turn will provide the student with the necessary skills that are required to make use of statistical knowledge in order to become an industrial practitioner.

Acknowledgement

I wish to thank my colleagues especially Mr. D.F. Phipps, for their help, comments and suggestions in the preparation of the paper.

References

[1] Bishop, H.E. (1964) The Training of Government Statisticians. Journal of the Royal Statistical Society, A, 127, 211-215.

[2] Bissell, A.F. (1975) The Use of Case Studies in Teaching Statistics. BIAS, 2, 29-32.

[3] Deane, M. (1964) The Training of Statisticians for Economics and Business. Journal of the Royal Statistical Society, A, 127, 216-218.

[4] Jeffers, J.N.R. (1976) Can Applied Statistics Be Taught? BIAS, 3, 17-29.

[5] Jones, J.D. and Kanji, G.K. (1980) The role of professional experience in statistical education. Statistician, 29, 196-203.

[6] Kanji, G.K. (1974) The role of the Statistical Laboratory in the Teaching of Statistics. Int.J.Math.Educ.Sci.Technol., 5, 53-57.

[7] Kanji, G.K. (1979) The role of projects in Statistical Education. Statistician, 28, 19-27.

[8] Kanji, G.K. (1982) Degree with a Difference. In: Proceedings of Concordia Conference on Applied Statistics, Marcel Dekker Inc. (To appear).

[9] Kanji, G.K. and Gilchrist, W.G. (1981) A Degree in Applied Statistics. Statistician, 30, 193-198.

[10] Oakes, M.B. (1981) Evaluating the Learning Outcomes. Proceedings of the Second World Conference on Co-operative Education, Northeastern University, Boston, USA, pp.142-146.

[11] Preparing Statisticians for Careers in Industry: Report of the ASA Section on Statistical Education Committee on Training of Statisticians for Industry (1980), American Statistician, 34, 65-75.

[12] Report of the International Statistical Institute Committee on the Integration of Statistics (1980) International Statistical Review, 48, 139-168.

[13] Spicer, C.C. (1964) The Training of Medical Statisticians. Journal of Royal Statistical Society, A, 127, 219-221.

[14] Yates, F. and Healy, M.J.R. (1964) How should we reform the teaching of statistics? Journal of the Royal Statistical Society, A, 127, 199-210.

TRAINING IN STATISTICAL QUALITY CONTROL

A F Bissell, Abergavenny

INTRODUCTION

The observations on SQC training in this introduction to the Workshop are based on my involvement, both statistically and technologically, in quality achievement and control over a period of over 30 years. My early technical experiences led me to recognise the important role statistics can play in quality operations, and eventually to specialise in statistics rather than in technology. Since that switch, I have been concerned with statistics (including SQC) in manufacturing industry; with the preparation of standards and other "official" documents incorporating statistical methods; with computer or calculator-based systems design for quality or quantity control; and with the training of both statisticians and industrial technologists in industrial applications of statistics.

RANGE OF APPLICATIONS OF SQC

Despite my declared interest in industry and technology, I wish to emphasise that SQC is not a topic of purely industrial application. If I may again quote my own experience, I have recently encountered important uses of what are often regarded as SQC techniques in such diverse fields (other than manufacturing industry) as:-

- Control of quality of information from tests in haematology laboratories;
- Monitoring of rates of occurrence of defects in ocean-going ships (for insurance purposes);
- Appraisal of results of social and other improvements in reducing congenital abnormalities;
- Retrospective analysis of accident and sickness rates;
- Control of quality of output of clerical and data preparation departments;
- Monitoring fuel consumption in buildings, vehicles, etc.

To these, many others could be added, including the monitoring of residuals from experiments or predictive models.

My object in making these remarks is that, in determining the target population for training in SQC, attention should not be confined to those concerned with industry in general, let alone with technology in particular. Some acquaintance with the fundamental philosophy and methods of SQC is useful to anyone who may encounter data collected in some kind of ordered sequence, or with using or designing procedures to discriminate between desirable and undesirable situations or operations.

TARGET POPULATION FOR SQC TRAINING

In the previous section, I emphasised the breadth of applications of SQC. I now turn to another dimension, that of the position of the recipient of the training on the hierarchal ladder of his organisation.

SQC training has tended to concentrate on those at the lower end of this hierarchy - with "SQC practitioners". My own experience, and that of many authors on the subject, is that the training of SQC practitioners without the active involvement of higher management is virtually useless. The success of certain industrial nations in penetrating the markets of Western Europe and the USA is attributable not only to their productivity (a field where the techniques of SQC can be usefully employed, incidentally) but to their top management interest in quality at all stages from design, through manufacture, to follow-up service.

It follows, therefore, that SQC should form a part of courses in business management, as well as in technology. At present, such business courses emphasise financial and marketing activities. The importance of quality, methods for investigating failure to achieve it, for assessing improvements and for monitoring performance, should be essential topics for all managers - the higher the current or potential status, the greater becomes the need for their interest, direction and inspiration in the area of quality.

Obviously the content of SQC training will vary according to the place of the individual in the organisation. The target population, nevertheless, will include production staff, designers, salesmen, market analysts, accountants, as well as the technical specialist. Many of these are likely, much to their surprise, to find that SQC offers insight and applications in their own areas of activity, in addition to those in connection with the products to which their specialism is related.

I need hardly add, as a rider, that a knowledge of the philosophy and methods of SQC is essential to the statistical consultant, and a course in SQC should have as high a status as one in regression (or experimental design, or multivariate analysis) in graduate statistics courses.

CONTENT OF SQC TRAINING

In view of my remarks on the target population for SQC training, and the consequent variety of needs, both in breadth and in depth, that this will generate, I can only offer here a list of topics which bear upon SQC. In order to promote discussion, I divide them into some broad categories. Thus, in considering how a trainee *statistician* might be catered for, we might reasonably assume his familiarity with the aspects listed under Basic Statistics and Statistical Techniques. On the other hand, the trainee *technologist* may need detailed training in these fundamentals. Again, the trainee *manager* will not require such a detailed training in the techniques listed under SQC and Reliability, but rather with principles and some examples of applications.

Basic Statistics	Data types and levels of measurement. Data presentation - summary statistics and graphical/tabular methods. Probability and distribution theory - with emphasis on *models*, not *laws.* Sampling methods.
Statistical Techniques	Inference, with estimation occupying at least as important a role as significance tests. ANOVA - including its use in estimation of variance components. Correlation - and its proper use in measuring association and dimensionality. Regression and linear models. Experimental design, with emphasis on the non-exclusivity of various designs, and including response surface representation.
SQC techniques	Sampling inspection (attributes and variables). Basis and procedures for control charts. Cusum methods - monitoring and retrospective analysis. Process capability measurement. Sample design for cost-effective control.
Reliability	Probability plotting (in my view, this might be better under basic statistics). Hazard functions and time to failure. Distribution models (Exponential, Gamma, Weibull, Lognormal, etc.) their applications and limitations. System reliability.
Other topics	National and International standards in quality, reliability and statistics. Use of statistical packages and systems. Simulation as a tool for Q. and R. study. Management practice (financial, personnel, technical, production, marketing) vis-a-vis QC attainment.

TEACHING AND TRAINING PRINCIPLES

The target population must be kept in mind when deciding in what manner and in what depth to teach those topics pertinent to the students or course members' needs. For training QC practitioners already working in industry, a much less theoretical approach is appropriate than when training statisticians for applied work in quality control.

Some general points that need to be noted are:-

We must strive to impart understanding of methods, not merely provide recipes.

Statistical methods are based on assumptions and principles which need justifying in the situations and data on which they are used. They also depend on probability and distribution models, which are not laws.

Methods of checking (calculations and assumptions) should be taught, and the importance of operating SQC procedures on regularly updated parameters needs stressing.

Impress on the QC practitioner that:-

Your data costs Time
Money
Effort

Collect it wisely
Analyse it effectively
Use it efficiently

In presenting topics to actual or potential QC practitioners especially if they lack a strong background in general statistical methods, a highly practical approach, with emphasis on visual aids, is appropriate. Many topics which cause difficulty when written as formulae or algorithms (e.g. probability rules, central limit theorem, components of variance, confidence interval estimation) lend themselves to dynamic visual aid presentation. A dash of humour is not amiss to liven up the subject.

TEXTS AND REFERENCES

In an introduction of this kind, it would be impossible to cover adequately the selection of texts and papers for use in SQC training. The choice will depend on the situation - general or "in-house" training? which industry? in what country? at what level? In any case, you will wish to express your preferences during the workshop discussions, and I will not therefore preempt or prejudice these discussions.

A few titles are obvious - general courses are likely to include (at least parts of) Juran's Quality Control Handbook, and perhaps a specific SQC text such as:-

Statistical Methods for Quality Control
Jardine, Mcfarlane & Greensted (Heineman)

Statistical Quality Control Methods
Burr (Marcel Dekker)

Sampling Inspection in SQC
Guenther (Griffin)

Sampling-Inspection & Quality Control
Wetherill (Chapman & Hall)

Appropriate national and international standards will be useful in many cases, and more general statistical texts in some, e.g.

Statistical Methods in Research and Production
ed. Davies and Goldsmith (Oliver & Boyd)

Statistics for Technology
Chatfield (Chapman & Hall)

Statistical Theory with Engineering Applications
Hald (Wiley)

At the appreciation level, "Facts from Figures" by M.J. Moroney (Pelican Books) provides a readable introduction covering a surprising amount of ground.

These are personal choices (at various levels of theoretical treatment) and others will no doubt appear during the workshop.

CONCLUSION

May I be permitted a personal fancy? Some 12 years ago, I was forced by a particular turn of events on a major project to take an interest in Markov Chains - a topic I had, till then, dismissed as of little relevance in my activities. Although the project was only vaguely connected with quality, I discovered I could then better appreciate certain papers in the literature which invoked Markov Chains in SQC techniques. The insight which this appreciation provided into the underlying unity of many SQC procedures, and the power of the Markov Chain representation as a tool for analysis of properties of SQC procedures, has convinced me that they should appear in courses for anyone who will be involved in SQC or Reliability theory, and any statistician who may be drawn into SQC or Reliability applications. With this possibly controversial suggestion, I conclude this introduction and offer the foregoing thoughts as a starting point for workshop discussion.

DISCUSSION OF PAPERS BY GREENFIELD, HAMAKER AND KANJI

A F Bissell, Abergavenny

The papers presented in this session deal with very different aspects of the training of industrial practitioners, but I detect three threads common to all three contributions. These are (not necessarily in order of importance) firstly that the industrial practitioner must be familiar with a broad spectrum of ideas, methods, techniques - some of them not generally included in conventional statistics courses, some quite outside the usual syllabus but nevertheless relevant. Secondly, statistics is an essentially practical subject, and this practical nature must be emphasised in training its practitioners. Thirdly, and closely linked to the previous point, the statistician must be able to communicate, at all stages of any project, with his clients, and the potential practician needs training in the various forms of communication currently available.

It is in the methods of imparting or cultivating these ideas that the speakers differ (though they do not fundamentally disagree). Professor Greenfield would teach a very wide variety of subject matter to young people about to embark on a career. Dr. Hamaker would, broadly speaking, take experienced practitioners in scientific disciplines and make statisticians of them (the route by which I came to statistics, incidentally). Dr. Kanji advocates learning by the experience of others, and by simulated experience, as part of a broad statistical training. The various approaches are not mutually exclusive. Each, whilst having the merits that the speakers have indicated, raises its own problems, and I now turn to consider each paper in order of presentation.

The list of subjects, topics, talents, techniques that Professor Greenfield presents is impressive, and one must agree that they are all relevant to the "compleate angler" trying to wrest all he can from the stream of data in which he wades. One might even add to the list - accounting methods, company law, human relations, etc. The danger seems to me that in order to squeeze these useful items into a degree course, we must sacrifice something else, either by omission, compression or superficiality. I wonder also whether the selection of (presumably) 18-20 year-olds for such a course can be effective - is this not too early to recognise potential "super-stats", let alone real polymaths? There would be problems at the end of their training too - we have already witnessed the failure (to reach the echelons of top management where the real decisions are made) of management trainees from business schools, and perhaps of O.R. scientists too. In an economic climate where only cash flow and return on capital are regarded as meaningful data, will the young polymath fare any better? Or will they, like their predecessors in numerate sciences, be dubbed with whatever is the topical derogatory term synonymous with "whizz-kid" or "boy-wonder", accused of having knowledge but not wisdom?

I agree that Britain, and other nations, needs polymaths, but is it perhaps the mature individuals, whom Dr. Hamaker recognises as potential applied statisticians, that we should be training in the way Professor Greenfield suggests? It would be essential to catch them at the right age - I suggest that they may not be sufficiently mature until 30-35, yet by 50 they will, with present industrial attitudes which seem likely to persist indefinitely, become candidates for so-called early retirement.

Dr. Hamaker sets less ambitious targets for his scientists or engineers. He aims at producing, or converting other practitioners into, good practical statisticians. I would question whether they can acquire an adequate statistical knowledge from text-books or extra-mural courses. Indeed, Dr. Hamaker's criticisms of some reputable text books could be construed as an implication that this is a route beset with dangers. A properly integrated course is essential to avoid the very dangers that Dr. Hamaker has pointed out. Statistics learned (or picked up) in an ad hoc fashion is likely to be compartmental, and will lack the depth of understanding necessary to avoid pitfalls like those in the examples quoted. Suitable structures already exist - the Institute of Statisticians examinations, for instance or the Sheffield Polytechnic course might be condensed to provide an M.Sc. or Diploma for mature students from indsutry. In the U.K., at least, this would raise problems of financing suitable candidates, and might need modifications to regulations in order to offer entry to those with the right experience but lacking formal first degree qualifications.

Dr. Hamaker, like Professor Greenfield, makes the point that statisticians should challenge the data they are presented with. I support and encourage such numerate scepticism, and would extend it to the questioning of foregone conclusions from previous investigations from which new projects often develop. For example, I recall being consulted about a specification which required that the ratio of two variables, say A/B, must be controlled to 1.25±0.04. This was based on some earlier work which purported to demonstrate how critical A/B was. Elsewhere in the same report was a record of subsidiary investigations from which it was evident that A and B could each be measured only to the nearest 2% - I leave you to consider the implication!

The first two papers dealt with what we may call the strategy of training industrial practitioners. Dr. Kanji has covered this elsewhere, not only in his papers in 1981 (with Gilchrist) and 1982, but in the actual development of the Sheffield Polytechnic course. In the present paper, he turns to tactics, and in particular the provision of vicarious or simulated experience. I agree whole-heartedly with the use of case studies, syndicate work and data generation equipment

03270

in statistical teaching at any level and for any field of application. The case study evidently needs to be oriented to the particular field of application, but equipment like the hot-wheel car can be used on any course. My version of the car-and-slope experiment is even simpler - a marble, a chute and a strip of carpet! One of the beauties of such demonstrations or data-gathering exercises is that the students can see the relevance of mathematical models in statistical work. In this experiment, if the car (or marble) is placed at the foot of the slope, its rolling distance is zero. Should we therefore constrain the constant term in the model to zero? And if so, is conventional least-squares the proper way to estimate b? I offer another example I use in teaching experimental design and analysis. Three participants tackle three jig-saw puzzles of differing sizes, timing their efforts, to assess whether the puzzles vary in difficulty. A Latin Square experiment takes only 15 minutes to execute, and provides data which illustrate, inter alia:-

(i) the use of a model to estimate effects and residuals;

(ii) the equivalence of a rule-of-thumb ANOVA to the analysis of a linear model;

(iii) the value of replication - three more participants, i.e. two replicates of the square, will quadruple the residual degrees of freedom and vastly improve the power of the experiment;

(iv) the dangers of using rule-of-thumb methods - in fact, a multiplicative model is more appropriate, leading to conventional analysis via log transformation;

(v) the necessity to re-interpret an analysis based on a transformation in terms of the original dimensions. Here, anti-logs of effects become "effect ratios" around a geometric mean.

Those interested in further examples of simple demonstration models will find a selection in "Statistical Teaching Aids" published by the Institute of Statisticians (Bury-St.-Edmunds, England).

I should like to take up a few points of detail in the three papers. Professor Greenfield's example of the restrictive view of the needs of statisticians in pharmaceuticals carries generally into the industrial field (and no doubt elsewhere). Indeed, the draft programme for this conference session lists SQC, Acceptance sampling and reliability as topics in industrial statistics:- my own experience is that there is (for the industrial statistics consultant) some of the first and third, little of the second, but much of the experimentation (not just experimental design), data sleuthing and effective data presentation.

I sympathise with Dr. Hamaker in his views that over-emphasis on "significance probability", rather than formal acceptance or rejection of hypotheses, is a move in the right direction - a procedure which some of us have been using furtively for years! I am less sure about the over-abundance of test methods. Certainly every statistician, industrial or otherwise, needs to know about robust and non-parametric methods.

Perhaps Dr. Hamaker is too emphatic about Bayesian methods. There is sometimes meaningful prior information which can strengthen inference about (for example) a regression coefficient or failure rate or mean life and I agree we must teach caution in the use of priors, especially when real data is sparse and the prior information may swamp it.

Do I detect, in Dr. Kanji's paper, a degree of disagreement with Professor Greenfield when he says that "it is not possible to train applied statisticians in the real sense at academic institutions"?. He makes this point although the Sheffield Polytechnic Course has an industrial experience layer in a 4-year sandwich (Professor Greenfield's is a 3-year non-sandwich course proposal).

All three papers give us much to consider, and (in the workshop sessions) to discuss and even argue about. I look forward to taking up in more detail one aspect of the training of industrial practitioners in Thursday's workshop on quality control.

DISCUSSION OF PAPERS BY GREENFIELD, HAMAKER AND KANJI

D G Watts, Queen's University at Kingston, Ontario

Mr. Chairman, speakers, and ladies and gentlemen. I would like to preface my discussion by thanking the Chairman for inviting me to participate in this session, and for organizing a very interesting and stimulating period. It is a pleasure to be here and to be involved in this important first International Conference on the Teaching of Statistics.

We have had the privilege of hearing three well-written and well presented papers today. Professor Hamaker's paper was concise, with an excellent summary, demonstrating the good communication skills recommended by all three speakers. He raised many good points, most of which I agree with, a few I didn't understand fully, and one with which I quite strongly disagree. Professor Greenfield's paper I found interesting and with some neat innovative ideas, but rather imprecise to serve as a blueprint for curriculum design. I will make some specific suggestions on that aspect later. Professor Kanji's paper was the most limited of the three, and offered some persuasive arguments for inclusion of projects in statistical training.

There were some obvious agreements of opinion in the three papers, the two most important to me I describe as

1) communication abilities, and
2) the statistician as scientist.

I strongly identify with both of these.

Some of you may remember a Conference on the Future of Statistics held in Madison, Wisconsin in 1967, the proceedings of which were published by Academic Press. One of the major themes of that conference turned out to be communication, not because of design, but because of its INTRINSIC IMPORTANCE TO THE DISCIPLINE. I commend those proceedings to you for some excellent reading on this important topic, which is now receiving a more appropriate share of attention. One thing you may be struck with is that not much progress has been made. Perhaps with enthusiasts like Professors Greenfield, Kanji, Hamaker and Snee and others, and hopefully influential reports like the ASA Section on Education reports on Training of Industrial and Government Statisticians, and the ISI report on Integration of Statistics, progress will be accelerated.

I suggest statisticians are going to continue to have trouble getting communication skills into the curriculum for a couple of reasons, however: many, if not most academic statisticians have no real experience in writing such reports, and therefore will resist - because it will be hard for them to come to grips with (e.g. marking is very time consuming, albeit rewarding) and because they don't appreciate the importance of these skills and therefore will not accord them a high enough priority in the curriculum: a manifestation of the "it can't be important if there is no equation on the board" attitude. To effect this change, then will require assistance from outside - from the consumers of our product - by demanding these skills.

One of the things that is being recognized (or discovered) by statisticians is that communication abilities are skills which can be learned, but that they must be learned by DOING. Report writing, presentations, and interpretations must be practised, ...and practised, ...and practised. Professor Greenfield suggests some nifty ideas to make this more enjoyable, through a weekly newspaper and a monthly magazine, and through a course conference. Another great way to encourage report writing and analysis interpretation skills is through projects and real experiences, as recognized by Professors Kanji and Hamaker, although I wish one could find more stimulating and relevant projects than the example of hotwheel regression or one we use at Queen's which measures rebound height of a dropped golfball. The golfball example at least has the virtue that a transformation of the response should be made to stabilize variance and subsequently the same transformation should be made to the drop-height to reinstate a straight line relationship.

One of the best ways to improve communication skills, in a nicely motivated way, is to have the students do their own experiments or to have them assist someone on the campus - projects, as recommended by all three speakers. I have had very good success in this, by having 3 or 4 students form a group and then find a researcher or graduate student on campus who needs to analyze some data. Each group must prepare PROPOSALS on at least two projects which I then use for evaluating the project as regards suitability for the course and viability. The students then use the project throughout the course to motivate them and as a learning agent. They must prepare a final report for submission to me for grading (35 - 50% of course mark) and to the researchers for their use.

The second issue which was strongly evident in all the papers was that of what I term the statistician as scientist, or what Professor Greenfield refers to as the polymath. One of the great advantages I believe my engineering training gave me over my statistical colleagues was that it took the STATIC out of statistics. That is, most statisticians are steeped in theory which assumes no dynamics - the independent part of I I D.

Three courses from my undergraduate program which I have found, and continue to find, particularly valuable, were very much concerned with problem solving. The first was a first year Civil Engineering tutorial course given in conjunction with the calculus course, in which we had to set up the differential equations for given physical situations, then solve the equations and finally write up the results in a workmanlike fashion. The second was a third year mathematical physics course, again where we developed mathematical models for various physical systems and solved the equations, very often using approximate methods. The third was a fourth year Electrical Engineering course dealing with the setting up of differential equations for dynamic systems, and again solving them. The emphasis was on SETTING UP, not solving. I have often felt that these three courses have been the most useful of my undergraduate courses, and this seems to concur with Professors Greenfield's and Hamaker's contentions.

Another advantage I believe my engineering undergraduate training gave me is something which is implicit in what Professors Greenfield and Hamaker say, and that is a different PERSPECTIVE from liberal arts students. Crudely, one can shoot the breeze with medical people or engineers more easily after having been faced with many of the same difficulties, which engineers refer to as "The perversity of inanimate objects". I also had to do many experiments and write them up.

I disagree with Professor Hamaker that Bayesian methods are to be avoided because "they make unrealistic assumptions". As an engineering statistician I find the Bayesian approach much more realistic and sensible than the sampling theory approach, and much more intellectually and scientifically honest. Bayes doesn't cause you to make unrealistic assumptions, it makes you ADMIT YOUR ASSUMPTIONS.

Which brings me to another perspective on perspective: one thing about being a scientist statistician (one identifying attribute being that you continue to consult and collaborate wherever you are) is that you are forced to adopt a decent perspective about assumptions and about analyses. For example, one realizes that many of the problems academic statisticians worry about are of about tertiary importance compared with the gross assumptions and simplifications one must make to formulate most engineering or medical problems.

I was a bit surprised about the comments on computing. Professors Hamaker and Kanji gave them no or little import, while Professor Greenfield recommended considerable programming skills in several languages. I disagree with all three speakers in this instance, because it seems to me that computer packages are becoming more useful and more powerful, and hence statisticians should not have to be adept in programming in several languages. We certainly need to be able to USE computers EFFECTIVELY, of course, so computing must be in the curriculum, but as "solution facilitators" I think we should be primarily computing MANAGERS, not computing PROGRAMMERS.

In conclusion, I would like to mention curriculum design, which Professors Greenfield and Kanji discuss. I believe that both these gentlemen and their respective Departments and students, - and we in other departments - would benefit if they would apply modern educational techniques of curriculum design in the style propounded by Robert Mager, especially in his book with Kenneth Beach Jr. "Developing Vocational Instruction", Fearon Publishers, California. The technique is very systematic and involves first writing a job description, which is "a general statement about what a person on the job does, and tells something about the conditions under which he does them". Next is the task analysis which requires "listing all the tasks that might be included in the job ..... and then the steps involved in each of the tasks on the list in terms of what the person does when performing the step". This is followed by development of prerequisites and course objectives, for example as applied to regression analysis in my paper "A Task-analysis Approach to Designing a Regression Analysis Course", The American Statistician, May 1981, pp.77-84. This approach is a marvelous one for clarifying exactly what should be IN a course as well as what should be OUT; it gives a wonderful operational perspective.

TOPIC 15: PRINCIPLES OF LEARNING PROBABILITY AND STATISTICS

ORGANISER: PROFESSOR E FISCHBEIN

LEARNING FROM EXPERIENCE - BAYES THEOREM: A MODEL FOR STOCHASTIC LEARNING SITUATIONS?

H Steinbring and G von Harten, University of Bielefeld

INTRODUCTION: THE BAYESIAN APPROACH - AN INTEGRATED BASIS FOR MULTIPLE DEMANDS?

One of the most difficult, and presumably most far-reaching, questions for stochastics education is that of the relationship between probability theory and statistics (Cf. Dörfler/Fischer 1981). From this perspective, it must be welcomed that at least one of the "Invited Paper Sessions and Workshops" within the frame of the First International Conference on the Teaching of Statistics will be devoted to the problem of "Principles of learning Probability and Statistics".

In his well-known book "Mathematics as an Educational Task", Freudenthal has warningly described the negative effects the separation and mutual dissociation between probability theory and statistics may have on mathematics education. Without the reference to the mathematical concepts and interpretations from probability theory, statistics can be reduced to a collection of cooking recipes. Cut off from its real applications, and practised purely within the framework of set theory, combinatorics, or measure theory, probability theory becomes a formal mathematical system. Summarising this comparison, Freudenthal draws the following conclusion: "We have given here a rough sketch of two kinds of teaching degeneration: probability as an abstract system cut off from reality, and on the other hand a system of patterns of computations to be filled out by numerical data. They look like opposites of each other but they are complementary in a most natural way!" (Freudenthal, 1973, p.591).

To oppose probability theory to statistics along the dividing line between mathematical theory and its applications also has a negative effect on the educational conceptions of teaching stochastics. Above all, this split can be rediscovered in the fundamental concept of stochastics, i.e. in the probability concept itself. The question is whether probability in mathematics instruction is to be introduced starting from applications, i.e. from statistics by means of the concept of relative frequency, or whether it is more appropriate to take the path via ideal gambles and thus via the classical conception which stresses the theoretical-combinatorial aspect of probability - a "wrong" alternative in our eyes.

This indissoluble tension between probability theory and statistics - "Probability theory is so close to its applications that it can even boast a theory how it is applied, I mean mathematical statistics". (Freudenthal, p.610) - has led, in the historical development of probability theory as well, again and again to debate and to attempts at founding this theory anew which from the very beginning strove to end this separation.

Thus, for example, the renowned probability theorist Richard von Mises attempted to reach this goal by means of axiomatising principles of experience, the so-called primordial phenomena (Urphänomene). Accordingly, he thought that the distinction was not to be made between probability theory and a theory of its application.

The Neo-Bayesian approach also pretends to formulate a unified point of view integrating both theoretical and empirical aspects which ends the separation between probability theory and statistics, or at least strongly relativates the latter. In the following, we shall discuss and analyse this point of view in more detail.

If we try to imagine such a point of view with regard to stochastics, we shall at once get into a situation which is typical, as a whole, for stochastics education. We are talking of a fundamental problem which is mainly relevant for mathematics instruction, and for the mathematics teacher: stochastics works with a variety of models (cf. Freudenthal's statement: "No scientist is as model-minded as the statistician."). Didactics, too, makes use of this promising model aspect. Just think of Engel's abacus, of the urn model, also of dice, coin-tossing and the wheel of fortune (which are objective representations of models). Other forms of representation, too, reflect aspects of stochastic models: attention may be called to tree diagrams, even set calculus, to other forms of graphs, and also the formulas used in connection with stochastics. Stochastics involves a whole range of most different types of modelling: starting from concrete models in objective form, across representations of these models getting ever more abstract, up to representing their mathematical form by means of fundamental equations and formulas. Within the context of such models and their various modes of representation, the attempt is to provide a most simple mathematical calculus, a basic and effective technique. As opposed to this technical simplicity of modelling, there are, however, the many problems of interpreting these models and of applying them to random situations. It is mainly this interpretation and application of these ideal models which constitute the fundamental problem: in the first place bridging the gap between the models, their forms of representation, graphs and formulas on the one hand, which are themselves at first given as fixed in their mathematical structure, and which basically function in a deterministic way, and real chance events on the other, concerns the aspect proper of stochastics, and may provoke stochastic thinking.

The mathematical calculus underlying the Bayesian approach is also of simple structure. Here again, the problem is not situated in mathematics, but in this model's interpretation, in the way it is applied. It is seen that such an approach which claims for itself tohave overcome the separation between probability theory and statistics is just as well affected by these problems in the same way, on a higher level. These issues concerning the relation between the mathematical form of representation and its interpretation, between simple theoretical models and their effective application, are, in the first place, crucial for learning stochastics, and for developing stochastic thinking.

The Bayesian approach shall serve us in this respect: which are the advantages and drawbacks of such a model, mainly with regard to a better understanding for stochastics education? This question refers to two different aspects: to both the exemplary use of such an approach in instruction, and to the benefits and more comprehensive knowledge resulting in principle from such a didactical way of phrasing the question.

Accordingly, and from an educational perspective, this is not simply a matter of deciding whether the Bayesian approach is right or wrong - i.e. of contributing to the controversy about the subjective and the objective concept of probability - but rather this debate is intended to help us obtain new insights into the relationship between probability theory and statistics which are not confined to stating either a strict separation, or an identity, between probability theory and statistics.

BAYES THEOREM AND "LEARNING THROUGH EXPERIENCE"

It is remarkable that the theorem from which the opposed groups of "Bayesians" and "Antibayesians" derive their names is rather simple from a mathematical point of view. The crux of this controversy is not about the question whether this theorem is mathematically correct, but about the problem of its application and interpretation.

Reverend Thomas Bayes had wanted to tackle the following problem: "Given the number of times in which an unknown event has happened and failed: Required the chance that the probability of its happening in a single trial lies somewhere between any two degrees of probability that can be named." (Bayes, 1763/1958, p.298).

His friend Mr. Price, who published Bayes manuscript after his death, thought that Bayes had mathematically solved the following problem: "Common sense is indeed sufficient to shew us that, from the observation of what has in former instances been the consequence of a certain cause or action, one may make a judgement what is likely to be the consequence of it another time ..." (Bayes, 1763/1958, p.296).

Loosely speaking, Bayes has solved the problem of estimating from an observation the probability of its causes. For a finite number of causes, Bayes' theorem has the following form:

If only the causes $E_1, E_2, \ldots, E_n$ are possible and if ω_i is the "a priori" (i.e. before observation) probability of the cause E_i, and further if p_i is the "a priori" probability of an event A in case the cause E_i is given, then the probability to be estimated $p_i = P(E_i|A)$ is proportional to $\omega_i \cdot p_i$, that is

$$(1) \quad P(E_i|A) = C \cdot P(E_i) \cdot P(A|E_i)$$

Summing over i yields :

$$C = \left[\sum_{i=1}^{n} P(E_i) \cdot P(A|E_i) \right]^{-1}$$

and hence this is Bayes' theorem :

$$(2) \quad P(E_i|A) = \frac{P(E_i) \cdot P(A|E_i)}{\sum_{i=1}^{n} P(E_i) \cdot P(A|E_i)}$$

Let us look at the following application:

(3) Given three urns with white and black balls. The probabilities of drawing a white ball out of one urn are $p_1 = 1$, $p_2 = \frac{2}{3}$ and $p_3 = \frac{1}{3}$, respectively. The urns are equally probable.
We have chosen an urn and drawn a ball: it is white. By (2) the probability that the ball belongs to the first urn is:

$$P_1 = \frac{1}{1 + \frac{2}{3} + \frac{1}{3}} = \frac{1}{2}$$

This application of Bayes' theorem is undisputed. The result can even be verified experimentally.

Let us look at another example:

(4) We choose at random n+1 points $P_0, P_1, \ldots, P_n$ in the interval [0,1]. The point P_0 chosen first has, like any other point, a uniform distribution. Knowing its position p, the event that any particular one of the other points $P_1, P_2, \ldots, P_n$ falls left of P_0 will have the probability p, independent of the others. If p is not known, the probability that k out of the other n points fall to the left of P_0 is $\frac{1}{n+1}$ for every k, because P is equally likely to be any one of the n+1 ordered points. If we assume that we know there to be k out of the n points to the left of P_0, then the probability of a success with the next point (i.e. that P_{n+1} falls to the left of P_0) chosen at random

is given by $\frac{k+1}{n+1}$, because the n+1 points divide the interval into n+2 pieces, k+1 of which are to the left of P_o. The probability distribution of P_o, which is initially uniform, is no longer such if we know that out of n random points k have fallen to the left of P_o. It is, instead, the beta distribution $f(p) = K.p^k(1-p)^{(n-k)}$, because P_o is then the (k+1)th point from the left of n+1 random points.

This example can be seen too as an application of Bayes' theorem, and the problem is very similar to another solved by Bayes in his article. What we require is a continuous version of Bayes' theorem:

Take an "a priori" distribution (for the sake of simplicity we take a distribution with a density) of the parameter p i.e. f (p). A is an observation. Then we have:

$$f_{\pi|A}(p) = \frac{f_{A|\pi}(p) \cdot f_{\pi}(p)}{\int_0^1 f_{A|\pi}(p) \cdot f_{\pi}(p) \quad dp}$$

If the values of p are "a priori" equally probable (i.e. f (p) = 1) we get:

$$f_{\pi|A}(p) = \frac{p^k \cdot (1-p)^{n-k}}{\int_0^1 p^k (1-p)^{n-k} \, dp}$$

In our example, we have to estimate the expectation i.e.:

$$E = \frac{\int_0^1 p^{k+1} \cdot (1-p)^{n-k} \, dp}{\int_0^1 p^k \cdot (1-p)^{n-k} \, dp} = \frac{k+1}{n+2}$$

In classical statistical analysis, the probability p, for instance, that a die will show a six, is viewed as a constant, unknown but fixed. Bayes' theorem now opens the possibility of a radically different approach, i.e. to consider the unknown value of the parameter p as a value of a random variable.

For the above problem, everybody will find this approach justified. In this instance, there is no controversy about the "a priori" distribution of the parameter p; it is given naturally in the problem.

But consider the following problem:

(5) I have a coin. I hold it in my hand and I am relatively convinced that this coin is fair. But I also toss the coin one hundred times. My "a priori" distribution (the density) of the probability of a head would, for example, have the following form:

$$f_{\pi}(p) = \frac{(2m+1)\,!}{(m!)^2} \cdot p^m \quad (1-p)^m$$

This distribution shows a maximum at $p = \frac{1}{2}$.

I toss the coin one hundred times and it turns up heads for $k = 48$ times. Now the probability for heads in the next toss (for $m = 10$) would be:

$$p = \frac{59}{121} = 0.4876$$

But if I had chosen the following "a priori" distribution

$$f\ (p) = \frac{101!}{99!} \quad p^1(1-p)^{99}$$

I get according to Bayes' theorem:

$$p = \frac{50}{202} = 0.2475$$

My probability distribution would get better if I continued to toss, but for 100 tosses, these two estimates are rather different.

Another important dispute arose about the class of problems where probability can be applied.

(6) Suppose you meet a stranger in a bar who offers to toss a coin to decide who shall pay for the drinks. He will pay if the coin falls tails, you will pay if it falls heads. The thought crosses your mind that the coin may be two-headed, but you are too delicate to ask to examine it. What is the probability of the coin being two-headed?

Some people may say that the question is meaningless, whereas a Bayesian needs only ask himself what is his personal "a priori" probability of the coin being two-headed.

This may be π; thus the probability of "coin is fair" is $1-\pi$.

Suppose the stranger tosses the coin and it falls heads. Now by Bayes' law, you have to adjust your probability in the following way:

$$P(\text{coin is two-headed}\,|\,\text{heads}) = \frac{\pi}{\pi \cdot 1 + (1-\pi)\cdot\frac{1}{2}} = \frac{2\pi}{1+\pi}$$

The following table shows how the probability of "coin is two-headed" after the toss (of heads) depends on your "a priori" probability:

prior	0	0.01	0.1	0.5	0.9	0.99	1
posterior	0	0.0198	0.0182	0.667	0.947	0.995	1

Note that if you are certain before the experiment that the coin is two-headed ($\pi=1$) or that the coin is fair ($\pi=0$), the experiment will not change your probability. This is generally true for any statement.

J. Berkson has pointed out the problem of the range of application of Bayes' theorem by means of the following example:

(7) A lady in the supermarket approaches the meat counter. "Here", says the meat man, "comes the lady who has bought 10 pounds of flour every Monday for the last five years; I'll bet she does it again". The lady has mistakenly put four instead of five 2-pound packages in the paper bag. The man places it on the scales which read 8 pounds. Trained as a Bayesian, he computes the posterior mean and charges the lady for 9.99 pounds.

Let us briefly summarise the two alternative views of probability i.e. "personal" or "subjective" and "frequentist" or "objective", not with the objective of deciding which one is "better", but in order to underline the difference between them.

One difference is that the subjectivist is willing to apply probability ideas to a wider class of problems, as we have seen in example (6).

Another problem in the subjectivist view is that once you have established your "a priori" distribution there is a kind of automatism, i.e. Bayes' theorem, that tells you how you have to change it. There is no place for an unexpected hypothesis. Subject and object are bound together into a knot, which, once established, never can be untied.

For the frequentist, on the other hand, probability is linked to the object. As Dinges put it "... the probability is given by the experimental design, it is an objective value ... For an objectivist, the statement 'the probability of A is p', where A is a well-described event and p is a number, is either true or false... There is no place for the notion of hypothesis." (Dinges, 1978, p.116).

So both views, we think, can be characterised by the relation they assign to subject and object.

INDEPENDENCE AND EXCHANGEABILITY

Independence is given in case the possibility of observed results influencing subsequent evaluations is specifically excluded. This says, for the subjectivist, that any possibility of "learning through experience" is excluded if the original opinion is based on independence, because the latter, by definition, requires that the original opinion will not be modified on the basis of any osbervation of results. If, on the other hand, you consider independence as a hypothesis, there are statistical tests that can support or reject this hypothesis.

From the subjectivist point of view, however, this problem is the motive for considering the notion of exchangeability. This is a condition for the common distribution of a finite or infinite number of random variables, just as independence is. Mathematically these notions have the following form:

A family $\{X_i\}$ of random variables is called independent, if for every
collection:

$$F_{X_1,\ldots,X_n} = F_{X_1}\cdot\ldots\cdot F_{X_n}$$

and is called exchangeable, if for every finite collection and every p

$$F_{X_1,\ldots,X_n} = F_{X_{\sigma(1)},\ldots,X_{\sigma(n)}}$$

where F denotes the distribution function.

If independent random variables are equally distributed, they are ex
But there are interesting examples of exchangeable variables which a
independent.

For example, suppose that an urn contains m balls, r of which have
written on them and m-r of which have a "O". If these balls are dr
from the urn, one at a time at random without replacement, and X_k
the digit on the kth ball drawn, then the sequence $X_1, \ldots, X_n$ is
able sequence; however, the random variables are clearly not indepe
With this notion, one can easily solve the following problem:

(8) In an urn, there are 50 balls: 49 black and 1 white. Two pers
alternately draw without replacement. The one who draws first the
ball has won. Does the first to begin or the second have the better
to win?

The variables are exchagneable. Therefore the chances to win are

We have thus seen that exchangeability, from a mathematical point of
is an interesting notion. It is, for instance, possible to prove the
law of large numbers for exchangeable sequences of random variable
Unfortunately, this is rarely done in text-books.

Let us briefly summarise some advantages of this notion:

a) As we have seen in example (8), it is a very natural notion (
two events A,B are exchangeable if P(A|B)=P(B|A)), which has a w
of application.

b) It can be shown that an infinite sequence of exchangeable var
is a probability mixture of independent, identical distributed variabl
This is de Finetti's well-known representation theorem. Thus this n
has a technical meaning for an objectivist, too.

We have thus two intuitive and mathematical concepts: independence
exchangeability.

One may wonder why a subjectivist denies the notion of independenc
First of all, independence, as we have mentioned, offers no possibil

learning through experience. On the other hand, de Finetti, for instance, argues that independence is a very vague and imprecise notion.

We, however, consider both notions offer the possibility of building a model of a real phenomenon. In this respect, they are links between theory and application, and are faced with the problem of theoretical terms versus real phenomena.

THE BAYESIAN APPROACH - ONE AMONG MANY IN MATHEMATICS INSTRUCTION

It is a fact certainly undisputed that the demands addressed to stochastics, as they have been formulated by the subjectivist school, have initiated positive changes in stochastics instruction as well. As compared to a widespread conception of probability as a rigid, objectivist concept which is quasi attached to the object of cognition, a more comprehensive view of the complex relationship between the subject and the object of cognition was increasingly developed. Stochastic activities cannot be simply confined to the aspect of the objectively observable and measurable; the 'subjective' side of arguing, estimating, and deciding must also be taken into account.

Such a conception of stochastics which, for the purposes of instruction, is expressed in a qualitatively new relationship between subject and object, between the learner and the subject matter to be learned, and which is usually not so explicitly thematized in mathematics instruction, has been formulated by Dinges in a general and fundamental way as a proposition: "I should like to propose to begin stochastics instruction always with decisions under [illegible]

[illegible]

[illegible]

be learned by means of stochastic concepts and methods, and what is learned, at the same time, about stochastics? We shall use this double question to analyse concretely the Bayesian approach to mathematics instruction.

Quite in keeping with Dinges' propositions, Riemer attempts, in his doctoral thesis (1985), to present an approach to stochastics which places the focus on the subjective concept of probability, and on Bayes' formula. The important fundamental idea is that certain (previous) assumptions, which are given subjectively and "a priori" by the students, be checked by means of an experiment in order to be finally revised. This basic idea of an "a priori" distribution and of modifying this distribution on the basis of experimental data is, for one thing, the prime concern of the Bayesian approach; in part, a promising educational aspect emerges here as well.

The following stages can be sketched for the stochastics course proposed: the introductory stage requires the students to make subjective estimates which are then revised by means of experiments and relative frequencies. The second stage then presents experiments which are to be repeated just once (resp. few times). In this connection, the Bayesian formula is developed via the inverted tree diagram. In a third stage, the Bayesian formula is applied several times in the sense of an iteration; computer simulation enables one to stage several hundred experiments and evaluations by means of the Bayesian formula.

On the whole, it must be noted that this approach via the Bayesian formula does not just refer, as textbooks usually do, to the technical and calculation-related aspects of this formula, but attempts, by repeatedly applying this formula, to develop the Bayesian [illegible] concept of learning through experience [illegible].

[illegible]

[illegible]

[illegible]

Bayesian cognitive situation itself be questioned?"

- The one-sided attachment of probability to the subject of cognition entirely loses from view the reference to the object. The prerequisites and conditions of application and usefulness of the Bayesian formula can only be discussed if there is a representation of stochastic's object of application and its specific particularities. The specific particularities of stochastic object express themselves in different aspects; they can be physical, biological, psychological, and social characteristics, but also simple symmetrical conditions. These object-related characteristics are introduced in the shape of previous assumptions, but are also aspects which are explored by means of stochastics.

One of the most basic and abstract expressions of stochastics' object-relatedness is stochastic independence. For the Bayesian approach, the lack of self-reflection of the conditions of application shows, among other things, in the fact that the concept of independence plays an unimportant role. "Subjectivists ... know independencies only as something very much derived, as extremes in a symmetrically dependent model. This inability to develop productively the independence concept is, in our eyes, a grave defect of the subjective approach." (Dinges, April 1981).

We shall use a basic concept of stochastics, the inverted tree diagram, which is also one of the proposed course described, as an example to explain our proposition.

The inverted tree diagram is used to conclude inductively from an effect (or several) to possible causes. The possible causes are assigned an "a priori" distribution, and additional information is used to draw conclusions as to the effect; [illegible]

[illegible]

concerns the question whether really all possible causes have been included to find out the effect, and second, it concerns the question as to the reference to the object in the context of the independence concept. "This methodology faces a number of serious problems. The most fundamental is that unknown or unsuspected failure mechanisms cannot be included in the analysis... Sometimes, the probabilities of breaching each safety barrier are not necessarily independent since "common-mode" failures can increase the likelihood of failure of one barrier once another has been penetrated. The probability of such common mode failure is uncertain unless the physical mechanism coupling the supposedly independent barriers is understood" (Nuclear Power, Issues and Choices, 1977, p.227).

Both objective and subjective approaches to stochastics should avoid one-sidedness and beware of setting this absolute. Rather, there should be an effort, for stochastics instruction, to develop multiple approaches in a parallel way, fruitfully referring to one another; there can be no unique and solely valid conception of stochastics. "Stochastics in school must organically link notions taken from various philosophical traditions, in particular

- stochastics as the mathematics of mass phenomena
- stochastics as the logic of uncertainty
- statistics as the technology of transforming data into insights
- stochastics as a theory of decisions." (Dinges, 1981, p.51).

References

[1] Bayes, Th. (1958) An essay towards solving a Problem in the Doctrine of Chances, Biometrika, 45, 296-315.

[2] Berkson, J. (1977) My Encounter with neo-Bayesianism, Internat.Stat. Review, 45, 1-8.

[3] Dinges, H. (1978) Schwierigkeiten mit der Bayesschen Regel, Math. Phys.Sem., XXV, 113-156.

[4] Dinges, H. (1980) Thesenpapier, Manuscript, Frankfurt.

[5] Dinges, H. (1981) Bemerkungen zur Unabhängigkeit, Manuscript, Frankfurt.

[6] Dinges, H. Gedanken zum Aufsatz des Rev. Th. Bayes (1702-1761), unpublished Manuscript.

[7] Dorfler, W. and Fischer, R. (1981) Stochastik im Schulunterricht, Wien.

[8] de Finetti, B. (1972) Probability, Induction and Statistics, Wiley & Sons.

[9] de Finetti, B. (1974) Theory of Probability, Vols. 1 and 2, Wiley & Sons.

[10] de Finetti, B. (1979) Probability and Exchangeability from a Subjective Point of View, Internat.Stat.Review, 47, 129-135.

[11] Freudenthal, H. (1973) Mathematics as an educational task, Riedel, Dordrecht.

[12] Hacking, I. (1965) Logic of Statistical Inference, Cambridge.

[13] Hacking, I. (1975) The Emergence of Probability, Cambridge.

[14] Hamaker, H.C. (1977) Subjective Probabilities and Exchangeability from an Objective Point of View, Internat.Stat.Review, 45, 223-231.

[15] Heath, D. and Sudderth, W. (1976) De Finetti's Theorem on Exchangeable Variables, American Stat., Vol.30, No.4 188-189.

[16] Hewitt, E. and Savage, L.J. (1956) Symmetric measures on Cartesian Products, Trans.AMS, Vol.80, 470-501.

[17] Hodges, J.L. and Lehmann, E.L. (1970) Elements of Finite Probability. San Francisco.

[18] Humphreys, P.W. (1977) Randomness, Independence and Hypothesis. Synthese, 36, 415-426.

[19] Jordan, K. (1972) Chapters on the Classical Calculus of Probability. Budapest.

[20] Lindley, D.V. and Phillips, L.D. (1976) Inference for a Bernoulli Process (a Bayesian view), American Stat., Vol.30, No.3, 112-119.

[21] Nuclear Power, Issues and Choices. (1977) Ballinger, Cambridge, Mass.

[22] Riemer, W. (1981) Heuristische Begründungen und Modelle zu zentralen Begriffen und Sätzen der Stochastik, Dissertation, Wuppertal.

CHILDREN'S CHOICE BEHAVIOUR IN PROBABILISTIC SITUATIONS

Ruma Falk, The Hebrew University, Jerusalem

Most decisions in uncertain situations involve comparisons of the probabilities of success under the alternative choices, and selection of that alternative where the chances of success are higher. Hence the ability to discriminate between probabilities of varying discrepancies is crucial in such tasks. The development of that capacity was examined in two experiments.

In a previous study (Falk, Falk and Levin, 1980), 61 children, from 4 to 11 years old were repeatedly presented with two sets, each containing blue and yellow elements. Each time, one colour was pointed out as the payoff colour (POC). The child had to choose the set from which he or she would draw at random a POC element in order to be rewarded. The sets were of varying size with different proportions of the two colours. Three kinds of materials were used: Pairs of urns with blue and yellow beads, pairs of roulettes divided into blue and yellow sectors, and pairs of spinning tops, likewise divided into two colours. It was found that roughly around the age of six, children started to select the higher of the two probabilities systematically. The dominant error, that characterized the choices of the youngest children and decreased with age, was selecting the set with the greater absolute number of POC elements. No other systematic erroneous strategies were unequivocally discovered.

However, the conclusions of that study were derived mainly from analysis of group results. The pooling of responses of children in an age group might obscure individual choice strategies. Furthermore, some indications of incorrect choice strategies, different from "choose the set with more elements of the POC", could not be followed up because of shortcomings of the experimental design, so that a definite determination of such strategies was not possible. For example, several children seemed to select the urn where there were absolutely less elements of the non payoff colour (NPOC), but the design included too few trials where choice according to that rule differed from the correct choice. Moreover, there was no differentiation between the pattern of correct choices and the one dictated by the strategy; "Choose the set where the difference between number of payoff elements and number of non-payoff elements is greater". Thus, the latter principle could not be diagnosed.

THE EXPERIMENTS

The purpose of the present study is to identify the principles of choice in individual response patterns.

Design. The design was similar to that of Falk, Falk and Levin (1980). In each trial, two urns were presented, each containing blue and yellow beads in different numbers and in different proportions. One colour was designated as the payoff colour (POC). The children had to choose the urn from which they would try to blindly draw a bead of the required colour. Success in drawing a bead of the POC granted the child a prize. Hence, in order to be maximally rewarded, the subject's problem was to identify the set with the higher probability of the desired colour. The set of trials was designed to allow differentiation between *relative* and *absolute* quantitative strategies. By "relative strategy" I mean the correct choice of the set with a higher proportion of the POC elements. An "absolute strategy" could mean either selection of the set with more POC elements, or selection of the set with less NPOC elements, or possibly even selection according to a larger (or smaller) total number of elements. Suppose, for example, that the urn on the left contains one blue bead and three yellow ones, and the urn on the right contains three blue beads and four yellow ones, and yellow is the POC. The correct choice, by a relative rule, would be that of the left-hand urn, since $3/4 > 4/7$, however, the absolute strategy: "select the urn with a greater number of POC beads" will result in choice of the right-hand urn, since $4 > 3$.

The trials of our previous study had been classified into three levels of difficulty (Falk et al., 1980): "Easy" problems, in which one of the two proportions is larger than ½ and the other smaller than ½; "Medium" problems, involving comparisons of ½ with a different proportion; and "Difficult" problems, in which both proportions are either larger or smaller than ½. The unique status of the fraction ½ is justified by the following: A child who is confronted with either an easy or a medium problem may compare *absolute quantities* of the two colours *within* each set and select that set in which the POC is a majority. That consideration would lead to the correct choice. Only with difficult problems is there no way to avoid comparing *ratios between* sets. Indeed, some of our subjects had verbalised such an absolute principle of choice: Sarah (9:9) (9 years and 9 months old) responded to an "easy" trial of the type (4 yellow, 6 blue) on the left *versus* (2 yellow, 1 blue) on the right, with blue = POC, by selecting the left set, and explained: "Here (on the left) there are more blues than yellows, whereas here (on the right) there are less blues than yellows". We found that level of difficulty did affect the rate of correct choices, the subset of "difficult" trials evoking the lowest rate.

Since less than half the trials in that study were "difficult", the inclusion of a majority of "easy" and "medium" trials might have partly accounted for the early age of emergence of the ability to discriminate between probabilities. In the present study, most of the trials are of the difficult category, i.e. presenting two probabilities on the same side of ½.

In all trials, in both experiments, the urn containing more beads of the POC, also contains more beads of the NPOC (relative to the other urn). Note that is is *impossible* to construct a comparison where the correct choice, i.e., the urn with the higher proportion of POC beads, will include less POC beads than the other urn and more NPOC beads. It is possible, of course, for the correct choice to include more beads of the POC than the incorrect choice and less beads of the NPOC. However, in such a trial both strategies of choice "more of the POC" and "less of the NPOC" converge on the same urn, without any conflict. Therefore, in order to be able to diagnose the particular strategy by the pattern of choices, the urn with more beads of the POC (and more beads of the NPOC) was always different from that with less beads of the NPOC.

In addition to that, the design enables identification of consistent choice of the set where the (absolute) *difference* between the number of elements of the POC and the number of elements of the NPOC is greater.

Experiment 1 comprised 32 trials controlling for all the above mentioned potential strategies.

Consider, for example, the following trial: (3b; 9y) vs. (5b; 10y) with b = POC (i.e., the left urn contains 3 blue beads and 9 yellow ones, and the right one contains 5 blue beads and 10 yellow ones, and blue is designated as POC). Obviously, the right-hand urn is the correct choice since $5/15 > 3/12$, however, the strategy of selecting the set with "less of the NPOC" will result in selection of the left urn, since $9 < 10$.

Let us now consider (1b; 2y) vs. (6b; 9y) with y = POC. The correct choice is the left urn, since $2/3 > 9/15$; however, the strategy of selecting the urn where the difference between the numbers of elements of the two colours in favour of the POC is greater, will lead to choosing the right-hand urn, since $9-6 > 2-1$.

All 32 trials were of the "difficult" category.

Care was taken to present all possible cross-combinations of trials that would turn up to be correct or incorrect by employing all potential strategies. Table 1 presents the distribution of all the trials (along with those of Experiment 2) by the quantitative characteristics of the correct urn, relative to the other urn. The first line in each cell gives the number of trials of that category and the second line gives an example of a trial of that kind. The examples always give the numbers of the two colours written in the same order in both sets, with the number of POC elements written first (irrespective of whether it was blue or yellow). The arrow points at the size of the subset of the POC in the correct choice.

TABLE 1

Classification of trials by relative magnitude of numbers of beads of both colours (POC and NPOC) on each side, and the difference in favour of the POC on each side.

Experiment 1

		Number of beads of both colours		
		Greater on correct side	Smaller on correct side	Total
Difference between number of beads of the POC and number of beads of the NPOC	Greater on correct side	4 (8 : 12) > (3 : 9) ↑	8 (4 : 1) > (5 : 3) ↑	12
	Smaller on correct side	4 (5 : 10) > (1 : 3) ↑	8 (5 : 1) > (12 : 6) ↑	12
	Total	8	16	24
	Conservation trials	8 (1 : 3) vs. (3 : 9) ;	(5 : 5) vs. (2 : 2)	8
			Total	32

Experiment 2

		Greater on correct side	Smaller on correct side	Total
Difference between number of beads of the POC and number of beads of the NPOC	Greater on correct side	5 (6 : 4) > (1 : 2) ↑	5 (2 : 1) > (4 : 6) ↑	10
	Smaller on correct side	3 (3 : 6) > (1 : 3) ↑	3 (3 : 1) > (6 : 3) ↑	6
	Equal on both sides	4 (2 : 3) > (1 : 2) ↑	4 (2 : 1) > (3 : 2) ↑	8
	Total	12	12	24

Note: Cells give the number of trials (upper line), and an example of a trial (below). The arrow indicates the number of POC beads in the correct choice. The second number is that of NPOC beads. The two numbers in the other parenthesis give, respectively, the same magnitudes in the incorrect choice.

The trial described in the upper right cell of Table 1, for example, presents a comparison between an urn with 4 POC beads and 1 NPOC bead, and an urn with 5 POC and 3 NPOC beads. The first urn (that in reality was assigned a right or left position as randomly determined) is the correct choice, since 4/5 > 5/8. Numbers of elements of both colours are smaller on the correct side (4 < 5 and 1 < 3), and the difference between number of elements of the POC and number of elements of the NPOC is greater on the correct choice (4-1 > 5-3). In contrast to that, in the trial described in the left cell of the second row of the table, that same difference is smaller on the correct choice, since 5-10 < 1-3.

Eight of the 32 trials were "conservation problems" presenting the same proportion of POC elements constructed of different absolute numbers of beads (an expanded fraction versus a reduced one). Two examples are given in Table 1. Obviously, there is no "correct choice" in conservation trials. These trials were introduced in order to test the prediction that a subject selecting the urn with "more of the POC" on non-conservation trials, will select the expanded of two equal fractions, and a child whose strategy is to select by "less of the NPOC" will choose the reduced fraction of the two.

Half of the non-conservation trials presented the correct choice on the subject's right side and half on the left. Likewise, the POC was evenly split between blue and yellow. All the features of the tasks were combined in a counter balanced and randomised design. The 32 trials of the experiment were divided into "Half A" and "Half B". The two halves were balanced with respect to all the variables of the design.

Experiment 2 consisted of 24 trials. The design was essentially the same as that of Experiment 1, precluding conservation items. See Table 1 for the distribution and for examples of trials.

A special feature of that experiment was that all trials came in pairs. The same two urns, with exactly the same two distributions of blue/yellow beads, were presented in the two trials of each pair (although their right-left location was randomly determined). Only the POC was blue in one trial and yellow in the other one (see Table 2 for examples of trials with their "mirror images"). Recall that the urn with the greater number of POC beads, always included also more of the NPOC beads. That way, the design enables to distinguish subjects who employ a relative strategy (the correct one) from those employing an absolute strategy. Consider, for example, the trial (6b; 4y) vs. (1b; 2y) with b = POC (see Table 1, first row of Experiment 2). One should select the urn (6b; 4y), since 6/10 > 1/3. If the POC is changed to yellow, the other urn becomes the correct choice, since 2/3 > 4/10. Such a switch of choices would signify the application of a relative strategy, relying on comparison of proportions. However, imagine a subject who systematically selects the urn with "more of the POC". He or she will select the (6b; 4y) urn in both trials of that pair, irrespective of the stated POC. Likewise, a child who selects the set with "less of the NPOC", will choose (1b; 2y) on both occasions. Such a pattern of results clearly violates the reversal of the direction of inequalities with change of sign (i.e., the rule stating that if $P_1 > P_2$, then $1-P_1 < 1-P_2$). A consistent pattern of correct choices will result in reversal of choices on all 12 pairs of trials, whereas a systematic reliance on an absolute strategy will yield 12 pairs of identical choices. Fluctuating strategies, or mere guesses, are expected to yield switches in about half the number of pairs. Thus, the number of pairs (out of 12) in which a subject chooses the same urn, will be indicative of the tendency of that subject to depend on an absolute strategy, although the specific nature of that strategy can be determined only by examining the choices. The two trials of a pair did not necessarily appear successively; their position was randomly determined among the 24 trials.

Experiment 2 was planned for younger subjects, on the average, than those of Experiment 1. Therefore, the length (and duration) of the experiment was shortened by omitting the eight conservation trials. Another change relative to Experiment 1, was that four of the 24 trials presented paired comparisons of probabilities on opposite sides of the probability ½ ("easy problems"). An example of a pair of such trials is given in the first row of Experiment 2 in Table 1. This was introduced in order to test the adherence of subjects to absolute strategies when carried to the extreme. The puzzling question was whether a child who consistently selects the urn with "more of the POC" (or "less of the NPOC") out of the two urns, on difficult items, will stick to that principle also when confronted with an easy trial. In the latter case one can resort to selection of the urn within which the POC is a majority, as in $(\underset{4}{2} : 1) > (4 : 6)$.

Subjects. Thirty five children took part in the research. Twenty in Experiment 1 and 15 in Experiment 2. These were mostly children of highly educated parents (most of the children were from Jerusalem and only five were Americans from Oregon). Five children of each of the ages 6, 7, 8 and 9 were tested in Experiment 1. Experiment 2 included five children in each of the ages 5, 6 and 7. The children were invited to participate in a "lottery game" and came voluntarily (and gladly).

Procedure. Each child was tested individually. On both experiments "Half A" was run first and "Half B" second, for half the children, and vice versa for the other half.

A preliminary lottery was performed three times, tossing a die and using candies as prizes for obtaining an outcome that had been stated beforehand by the child. The purpose of this procedure was to acquaint the child with the idea that a win in a lottery is conditioned on the outcome of an uncertain event, and to prepare the child for possible disappointments.

Two experimenters were present. One gave instructions and presented the materials, and the other one recorded the responses.

One preliminary demonstration-trial with urns and beads preceded the experiment proper. Then came the two halves of each experiment with a short intermission for rest in between.

The standard procedure of each trial was as follows: The child was offered a selection of 3 to 4 prizes such as coloured candies, plastic toys or marbles. He or she selected the most attractive prize to try to win. The prize, together with a bead of the POC of that trial, were placed in front of the child in a small plate. The pair of urns were now presented and placed on both sides of the prize in equal distances. The beads of the two colours were put into the urns, according to the required composition, while the child was watching, then the experimenter explained: "You will soon be given a chance to draw one bead, with your eyes closed, from one of these urns. If you draw a blue (yellow) bead (indicating the bead next to the prize), you will win the prize. Which of the urns would you like to try to draw from?". The child then pointed at the selected urn, and closed his/her eyes. The selected urn was shaken by the experimenter and presented to the child's grasping hand. At the end of each trial the child either won or did not win the prize, according to the rules of the game. No verbal reinforcements were given for correct choices.

Verbal responses were not required of the child throughout the experiment. However, each spontaneous comment or explanation of the child was recorded. A short post-experimental interview was conducted with each subject. A couple of trials that had appeared in the experiment were presented again and the child was asked to explain his/her choice.

The experiment lasted about an hour (somewhat less for Experiment 2).

RESULTS AND CONCLUSIONS

All the responses (to non-conservation trials) were scored by either + or -, depending on whether the higher or the lower of the two probabilities had been selected, respectively.

The overall rate of correct responses, for the different ages, is presented in Table 2. The results describe both experiments, and are pooled across subjects and across all categories of trials.

Clearly, there is improvement in performance with age. The rates of correct responses are lower than in our previous study for all ages (see Falk et al., 1980, Table III, page 192). This was expected, considering that a much higher proportion of trials in the present study are "difficult", and that all trials present more beads of the two colours in the same urn. The conjunction of these two factors probably caused the performance in this research to be lower than in our previous study, even for the subset of trials with smaller number of POC elements on the correct side (Falk et al., 1980, Table IV), and for the subset of "difficult" trials (Table V). However, the overall rate of correct responses presents group results, eventually including individual subjects with contrasting response tendencies.

TABLE 2

Analysis of Results of Experiments 1 and 2 by Age of Subjects

			Proportion (b) of subjects deemed to:-			
Age Group	Number of subjects	Rate (a) of correct responses to all trials	master the task	(1) select by "more of the POC"	(2) select by "less of the POC"	(3) select by "greater difference in favour of POC"
5 - 6	5	0.56	-	-	0.40	-
6 - 7	10	0.57	0.10	0.40	0.10	-
7 - 8	10	0.73	0.50	0.10	-	0.20
8 - 9	5	0.78	0.60	-	0.20	0.20
9 - 10	5	0.86	0.80	-	-	-
Total	35	0.68	0.37	0.14	0.11	0.09

(a) Rates were computed out of 120 responses when 5 subjects were involved, and 240 responses in the case of 10 subjects.
(b) Proportions were calculated out of the number of subjects in the age group.

Each individual subject's pattern of choices was analysed in order to determine whether that subject had mastered the task of selecting the higher of two probabilities, and in order to diagnose possible strategies of choice.

The criterion for classifying a subject as employing the "correct principle", was that he/she achieved at least 18 correct choices out of 24 (non-conservation) comparisons. This means that no more than six errors were tolerated. Such a high level of performance would have been obtained by chance (were the child merely guessing) with probability 0.011. The same (arbitrary) level of proximity was also applied for determining that a response pattern agrees with the pattern predicted by any of the strategies studied. (There was only one exception: As can be seen in Table 1, there are only 16 definite predictions of responses resulting from the strategy: "select the urn with a greater difference between number of POC beads and number of NPOC beads", in Experiment 2. In that experiment the difference between the numbers of beads of the two colours is equal in both urns, in 8 trials. Therefore, the criterion for determining application of that strategy, was at least 13 responses, out of 16, matching the strategy. The probability of satisfying the criterion, by chance, is 0.011).

Table 2 presents the proportion of subjects, in each age, that either performed correctly (close enough to perfection), or applied one of the three strategies consistently. The three strategies that were "predicted" by the design, and did appear in the experiments, were: select by (1) "more of the POC", (2) "less of the NPOC", and (3) "a greater difference in favour of the POC". Note, that when the unit of analysis changes from trial to subject, numbers become small. The proportions in Table 2 were based on 10 or 5 subjects. Thirteen of the 35 subjects applied the correct principle, whereas 12 were sticking to different incorrect strategies: 5 subjects to strategy (1), 4 to (2), and 3 to (3). The mean number of errors for the 13 subjects deemed to master the task was 2.8. The mean number of deviations from a perfect "strategy-pattern", for subjects classified as applying specific erroneous strategies was 2.0.

A count of the number of identical choices in "symmetric" pairs of trials (i.e. pairs of trials presenting the same two urns with blue as the POC in one and yellow in the other) in Experiment 2, turned out as follows: There were 5 subjects who applied an absolute strategy (either "more of the POC" or "less of the NPOC"). These subjects made, on the average, 10.2 identical choices out of 12 pairs(1). All other subjects of that experiment, who did not apply an absolute strategy, scored a mean of 5.6 identical choices. Thus, the number of choices of the same urn, in spite of change of the POC, supports the impression that some of the subjects were committed to their erroneous absolute strategies. Furthermore, these 5 subjects stuck to their erroneous absolute strategies even on "easy" trials, that served as test-trials in Experiment 2.

Some individual results. A few of the children who matched the pattern of responses predicted by an absolute strategy, seemed very dedicated to their principle. Thus, Tali (6:0), in Experiment 1, who applied the strategy "more of the POC" in 23 out of the 24 non-conservation trials, verbalized her rule of choice explicitly in the interview: "Here (on the left) there are more blues (the POC of that trial), so I thought it would be easier for me to find a blue one". When we remarked that on the left there are also more yellow beads, and asked whether she had considered that, she replied: "That does not matter at all!". On one occasion, when we started filling the urns for a certain trial, and had already put the required numbers of POC beads in both urns, she immediately selected the urn where we put more beads. When told: "But we did not put the beads of the other colour, yet" she said: "This does not make any difference!". Not surprisingly, on all 5 conservation trials, Tali selected the urn with "more of the POC", i.e. the expanded fraction.

---

(1) The probability of making by chance at least 10 identical choices, out of 12, is 0.019.

Beni (6:0) also selected by "more of the POC" and consistently selected the "expanded urn" on all conservation trials. Yotam (8:1) applied the strategy of selecting "less of the NPOC". He too carried that pattern of choices over to conservation trials, where he chose the urn with less beads of both colours on 7 trials out of the 8. Ran's (8:6) principle was to select the urn with "greater difference in favour of the POC", and he verbalized that principle clearly. In 4 conservation trials, where the probability was smaller than ½, Ran stuck to his principle. Thus, in (2b; 6y) vs. (1b; 3y) with b = POC, for example, he selected the right-hand urn, since 1-3 > 2-6. (In the other 4 conservation trials the probability was ½ and choice by difference was impossible).

Among other children, who were not committed to any absolute strategy, all the range of the distribution of responses to conservation trials could be found. Curiously enough, a few of the children who responded correctly (beyond criterion), chose the reduced urn consistently on all 8 conservation trials. A refreshing reaction was that of Oren (9:7): When he was confronted with conservation problems, after coping with paired comparisons of different probabilities, he burst out in laughter, and contemptuously selected via chanting "En Den Dino" (a well known Israeli nursery rhyme for chance selection of one from a group). The procedure proved efficient... his responses were equally divided between the reduced and the expanded fractions.

Another absolute strategy, that appeared in Experiment 2, deserves mentioning: Tobi (5:3), who was classified (in Table 2) as choosing by "less of the NPOC", was actually selecting the urn with a smaller total number of elements. Since it was the same urn, on all trials, that contained less of the POC and less of the NPOC, that urn also had a smaller total; consequently there was no differentiation between the patterns of choices resulting from the two strategies. However, Tobi counted the beads aloud, in the course of the experiment, and he was obviously comparing totals, and he said so clearly in the post-experimental interview. His adherence to the strategy was perfectly consistent on all 24 trials, even in the "easy" trials of that experiment. Tobi's number of identical choices in the symmetric pairs of Experiment 2 was 12, i.e., in all pairs of trials he chose the same urn, in spite of reversal of POC.

A "dramatic" insight, in the course of the Experiment, was displayed by Shahar (6:3). He started by selecting "more of the POC", and did so consistently on all the trials of "Half A" and on the first 4 trials of "Half B" (he was therefore classified among the subjects applying that strategy in Table 2). At that point he announced spontaneously: "Now I shall choose where there are fewer" (probably meaning: "fewer of the NPOC"). He did switch to that strategy for a few trials, but later abandoned it and kept fluctuating. The point of interest for us was watching him suddenly discover "the other colour", and later on exhibiting some conflict and lack of knowledge of how to utilize the information on both colours simultaneously.

Naomi (7:4) spread her 7 errors over 7 different symmetric pairs of trials. She was focusing on the same urn in these seven pairs (and consequently made identical choices on 7 of the 12 pairs). She expressed her special interest in the urn when at one point of the experiment, after displaying some conflict, she said: "I am much attracted to this urn".

The issue of stages. Not all the subjects could be classified as selecting according to a given strategy, correct or incorrect. There were 10 children (of 35) whose "stage" could not be determined even with a less than perfect criterion of agreement. Some children seem not to stick to a single strategy, They either have some rule in mind and make errors in using it, or they fluctuate between strategies, or even make chance decisions (at least from the point of view of our analysis).

The conclusions that can be drawn from the results, concerning the developmental course of the ability to discriminate between probabilities, are limited. I shall try to delineate a tentative developmental course based on logical analysis, and partly on suggestions derived from the results.

It seems reasonable that absolute quantities would initially be dominant in that kind of judgement. Especially, the absolute size of the winning category (POC) would be salient in capturing the judge's attention. Hence, "select by more of the POC" would appear to be a candidate for the first quantitative strategy. Later, would come the understanding that the size of the set of elements of the other colour is also related to the probability of success. The awareness of the relevance of the number of NPOC elements may be dominant for a while and the child may focus on that number exclusively. That way, "less of the NPOC" would be the "second stage". The reliance on number of elements of only one colour may fluctuate between the two colours for a while, until the conflict is resolved by a realisation of the need to consider both quantities simultaneously. The integration of these two numbers via computation of proportion (or ratio) is not expected to occur instantaneously. Different attempts to integrate these two numbers may appear, combined with temporary "regressions" into judgement by only one of the quantities. One of the easiest and tempting ways to try to combine the two numbers, is via computing differences. Additive operations come first in education and hence are easily available to the child. This suggests that "select the urn with a greater difference in favour of the POC" could be the third stage, one before finally combining the numbers via proportion and mastering the conservation of that concept.

However, there is not enough evidence in the results to establish that sequence of strategies as an empirical finding. More data should be collected in order to learn about the prevalence of the various strategies in different age groups. Also a longitudinal study, following the same children for a number of years, may determine whether those stages appear sequentially, and whether they come necessarily in the same order.

Concluding remarks. One of the educational implications of this study is that the teaching of probability to children should be carefully planned. The lesson to be learned from the results of the present research, in conjunction with earlier studies, is that the concept of proportion (and probability) is very elusive. Children can easily avoid directly coping with proportional computation. They may find alternative strategies and appear as if they have mastered the problem. Some of the incorrect choice strategies that were diagnosed in the elaborated design of the two experiments, were not and could not have been discovered in some of the previous designs. For example, Yost, Siegel and Andrews (1962) presented paired comparisons only between complementary probabilities. Two unequal complementary probabilities are always on different sides of ½. As we saw above, such "easy" problems can be correctly solved by absolute considerations within an urn, whereas the experimenter, and teacher, may entertain the illusion that the children have mastered the task. Yost et al. (1962), Davies (1965), and Goldberg (1966) concluded that preschool children (even 4 year olds) were capable of systematic probability judgements. However, their comparisons always included an equal number of elements of one of the colours, the colour preferred by the child, in both sets. Hence, when the NPOC appeared in equal numbers the child could choose by "more of the POC", and when the POC came in equal numbers the child could choose by "less of the NPOC". That child would be counted as "capable of systematic probability judgement". In a similar vein, the few children who were judged in the present study as choosing by "a greater difference in favour of the POC" would have been counted as selecting correctly in Falk et al. (1980), because of the limitations of their design. We should take care that this would not happen in the teaching of probability. The teacher's responsibility should be to present the concept in its most general form and to make sure that all the components of the concept (namely, the numbers of elements of both categories, the difference between them, and the total) will appear in all their possible mutual relations and will vary in a free interplay.

Comparison of probabilities is but one of the judgements involved in choices in probabilistic situations. Decision making under uncertainty requires also assessing the costs and payoffs associated with the alternative choices.

Next comes the need to integrate the probability and the payoff (or utility) of each alternative into one magnitude characterising its expected value. A natural extension of the present research would confront the child, on each trial, with a task of choosing between two gambles. Each gamble would consist of its own probability of success and the payoff (gain) for winning. A conflict could be created between the result of comparing the two probabilities to that of comparing the payoffs, so as to diagnose possible choice strategies of children. Such a research is now under way.

Acknowledgements

This study was supported by the Human Development Center, The Hebrew University of Jerusalem.

Thanks are due to my husband Raphael Falk for his helpful comments and to my son Oren for assistance in conducting the experiments.

References

[1] Davies, C.M. (1965) Development of the probability concept in children. Child Development, 36, 779-788.

[2] Falk, R., Falk, R. and Levin, I. (1980) A potential for learning probability in young children. Educational Studies in Mathematics, 11, 181-204.

[3] Goldberg, S. (1966) Probability judgements by preschool children: Task conditions and performance. Child Development, 37, 157-167.

[4] Yost, P.A., Siegel, A.E. and Andrews, J.M. (1962) Nonverbal probability judgements by young children. Child Development, 33, 769-780.

AN IDEA OF STRUCTURING PROBABILITY TEACHING BASED ON DIENES'S SIX STAGES

Anna Olecka, Guardamangia, Malta

INTRODUCTION

Zoltan P. Dienes has proposed a way of developing mathematical concepts going step by step from concrete to formal operations. He has subdivided this process into six consecutive stages. These [1] are briefly:

(1) Initial interaction.
(2) Discovery of regularities in situations.
(3) The comparison of several games possessing the same structure; this is the stage of searching for isomorphisms.
(4) The representation of isomorphic situations in one all-embracing form.
(5) The study of the representation by the description of its properties. In this way all the isomorphic situations are being studied. Usually symbols will be needed at this stage.
(6) The formalisation of the system.

The purpose of this paper is to suggest how Dienes' idea can be applied as a framework for organising probability teaching. If we try to organise probability teaching globally in this manner, it turns out too complicated and inadequate; too many concepts are involved at the same time. So I believe that for the sake of didactical analysis it may be useful to separate particular concepts and to model the teaching according to Dienes' stages - each concept separately - and finally combine these models.

This approach may be a starting point for designing probability curricula for beginners: children, adolescents and even college students and teachers (if they are learning probability for the first time). Of course, young children will need more time for experimentation. Also the time allotted for the choice of apparatus, use of graphical schemes etc. would depend on the kind of student for whom it is intended. But the general pattern will follow the same line. And it is just this general pattern which is presented here. Neither methods nor teaching aids presented in this paper are new. They have been widely discussed in the literature for the last few years. I have taken advantage of them to fill up the idea of Dienes' stages with some practical contents. The literature on which this paper is based is listed at the end.

The concepts considered in this paper are: probability multistage experiments, expectation and conditional probability. However, only stages I - V will be discussed here. Stage VI could be common to all of them as a formal presentation of an elementary probability course in a systematic way, after completing stage V for the concepts involved.

THE CONCEPT OF PROBABILITY

Stage I. Initial interaction. Random experiments with various devices (coins, spinners etc.). Experiments may be connected with games, but the games should have very simple rules.

Stage II. Discovering regularities. The aim of this stage is to help the children to observe the stabilisation of the relative frequencies when the number of trials is large enough and to compare the frequencies with the chances assigned "a priori", intuitively or with some reasoning.

Two kinds of students' activities are expected at this stage:

- Experimentation, recording and organising data, computing the relative frequency, presenting through graphs both the data and the relative frequency.
- Some simple reasoning involving symmetry or proportions, leading to the theoretical values of probabilities in very simple cases (one- or two-stage experiments). Some graphical aids (tables, tree, diagrams, etc.) are useful here. In order to present the random experiment from the theoretical point of view, theoretical graphs (the different outcomes on one of the axes, the probabilities on the other) can be made.

Comparing theoretical chances with those otbained from estimation of the relative frequences is a significant, indispensable step towards understanding probability. It can be done by comparison between theoretical and statistical graphs. For some experiments it is impossible or at least very hard to assign theoretical probabilities of outcomes. Yet it is important to show some situations of this kind and to try to estimate probabilities by relative frequencies. A cumulative relative frequency graph will be a very useful tool at this point.

Stage III Searching for isomorphisms. During stage II, students carry out research on each experiment separately. In the next stage, it is time to make comparisons between the different experiments.

The aim of stage III is to show that some random experiments can be replaced by others.

At the beginning we shall consider cases where the random devices are isomorphic in some obvious sense. For example it is significant that students can grasp the idea that a spinner (Fig.1) can be used instead of a coin. This will not be so obvious if we divide the spinners into more than two parts (see Fig.2).

FIG.1

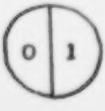


FIG.2

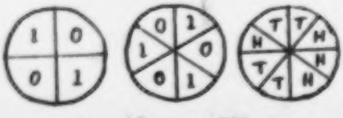
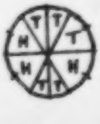


FIG.3



If we divide the spinner into unequal parts (as in Fig.3) it would be even more difficult. A comparison between theoretical graphs of the experiments yields a good criterion for isomorphism of the random experiments involved. When it is intuitively understood by students we can go on to the difficult area of comparing experiments with quite different structures. Example:

A. Spinning the spinner as in Fig.4 twice and recording the sum of the two numbers obtained.

B. Spinning the spinner as in Fig.5 once and recording the number obtained.

FIG.4

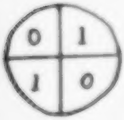


FIG.5

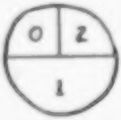
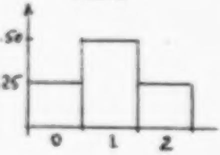


FIG.6



Although the experiments A and B are different their sets of outcomes and the corresponding probabilities are the same. Students can discover this while constructing theoretical diagrams for both of the experiments (Fig.6). Also circle diagrams for both of the experiments are the same: this is a spinner as in Fig.5.

This stage seems to be the crucial point in the learning process. So we have to choose adequate activities which would enable the students to simulate some probabilistic situations with others.

Stage IV. Representation. In order to describe a random experiment one has to describe the set of all possible outcomes and then to assign the probabilities of the outcomes. Thus the following table is a very useful representation:

outcome	w_1	w_2	w_3	...	w_n
probability	p_1	p_2	p_3	...	p_n

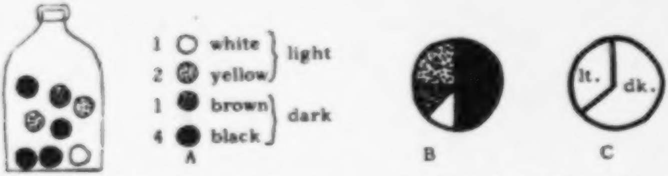
$p_1 \le p_2 \le \ldots \le p_n$

A circle diagram (Fig.7) or a one-step tree diagram can be another suitable representation. The circle diagram helps to see that $p_1+p_2+\ldots+p_n = 1$, as we just divide the whole circle into parts. The tree diagram will become very useful later on, when multistage experiments are considered.

Stage V. The properties of the representation. The main property to be discovered now is a rule for computing the probabilities of events. Sometimes one can do it just by changing the way the set of outcomes is described. For example, both of the circle-diagrams (as in Fig.7) represent an experiment of drawing a marble. The set of given marbles is the same in both cases

(Fig.7A). But in the first case (Fig.7B) the colour is recorded whereas in the other case (Fig.7C) only two possibilities are considered: "dark" or "light".

FIG.7



One can notice that we can get the same result by simply adding the probabilities of corresponding outcomes in the first case. So we get a rule for assigning probability to an event E:

- to describe (in any convenient way) all the possible outcomes and to assign them probabilities;
- to add the probabilities of all the outcomes favourable to E.

The classical scheme (when all the outcomes are equally likely) will appear at this stage as a particular, but significant, case.

Now some properties of probability may be studied (such as probability of the complementary event etc.).

MULTISTAGE EXPERIMENTS

Stage I. Initial interaction. At this stage the students will be doing and watching various multistage experiments (perhaps connected with simple games).

Stage II. Discovering regularities. Each experiment will now be separately investigated. The basic skill developed here is the ability to distinguish successive stages of a multistage experiment. In order to describe such an experiment the student will need some graphical representation. Tree diagrams are the most convenient, but for the two-stage experiments square diagrams or just simple tables are also useful.

Examples:

A. Drawing balls from the box ⌊○ ● ● ●⌋ without replacement until a white ball is drawn. A tree digram describing the experiment is shown in Fig.8.

B. Drawing two balls from the box ⌊○ ● ● ●⌋ without replacement. Constructing a square diagram describing the experiment is shown in Fig.9.

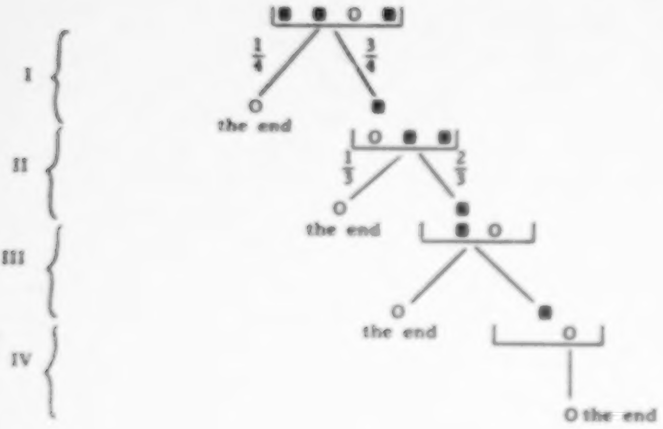
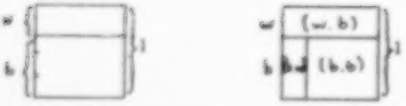


FIG.8



first draw

second draw

FIG.9

We assume that students are already able to assign correct probabilities at each stage separately, but computing probabilities for the entire experiment would be too difficult at the moment. So they would only be able to estimate them by repeating the experiment many times. A tree diagram can be helpful: we put some chips at the top of the tree and then each chip will follow a pattern down the branches according to the results of the experimentation. Finally we just count how many chips get to the end of each branch, and calculate the relative frequencies (number of chips at the end of a branch divided by the total number of chips used).

Stage III. Searching for isomorphisms. Here we shall compare various experiments. We can do it either by comparing all the possibilities they yield or by comparing their tree diagrams.

Stage IV. Representation. Two experiments will be considered isomorphic if they can be described by the same tree diagram (up to permutation of branches). Therefore a tree diagram could be a good representation of a multistage experiment.

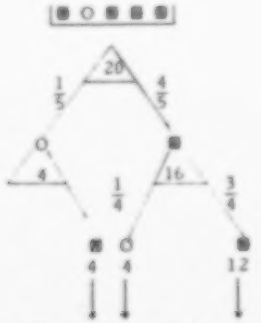
Now the question arises how to compute the probabilities of outcomes. One can use the special case of the probabilistic abacus of Arthur Engel [4], namely with the tree diagram. The probabilistic abacus is a kind of idealized simulation. The chips are moved along the branches and they are divided into groups, at each stage according to the probabilities of branches. Eventually each of them reaches the end of the branches (we supply more chips at the start if necessary). We can now define the probability of the whole branch as the fraction: number of chips at the end of the branch divided by the total number of chips used.

Using the probabilistic abacus on the tree diagrams may seem trivial, but it can be a significant step between estimating probabilities and computing them.

Finally, a natural representation of the multistage experiment is the tree diagram supplied with the probabilities of the outcomes. Fig.10 shows the representation of the experiment of drawing two balls from the box |● ○ ● ● ●| without replacement while two possibilities are considered: "the two balls drawn have the same colour", "each of them has a different colour".

Stage V. The properties of the representation. The above way of computing probabilities will lead to discovering the two fundamental rules: the rule of multilying probabilities along a branch and the rule of adding the probabilities of the favourable branches (see [2]). Later on we can examine carefully the tree diagrams in order to find properties of probabilities in multistage experiments.

FIG.10



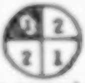
outcome	different colours	the same colour
probability	$\frac{4+4}{20} = 0.4$	$\frac{12}{20} = 0.6$

Stage I and stage II to some extent can be done together with multistage experiments. Indeed in describing the experiments with tree diagrams one has to assign conditional probabilities to the branches. Therefore we may regard these activities as the intuitive base for conditional probability.

Stage II. Discovering regularities (continuation). Our starting point was that the concept of probability should develop as a synthesis of both of the approaches: "a priori" and "a posteriori". The same idea governs our approach to conditional probability. Example:

Take the spinner shown in Fig.11 and record in the table below the numbers obtained, taking into account at the same time the colours of the fields obtained:

FIG.11



white	2	2	1	1	2	1	1	2	2
not white	1	1	1	1					

After having recorded the data for, say, 100 trials, we now disregard the lower part of the table and we calculate the relative frequencies of numbers 1 and 2 for those times when the spinner landed on the white field.

Another example:

Toss a coin 3 times and consider the event A: "we get an odd number of tails". Let us record the results of experimentation in the following manner:

A	(T,H,H), (H,H,T), (T,H,H), (H,T,H), (T,T,T), (H,H,T), ...
A'	(T,T,H), (T,H,T), (H,H,H), (T,T,H), ...

Now we consider another event B: "we get one tail only". We shall find the relative frequencies of B (i) from the combined results from both parts of the table (ii) from the upper and lower parts separately (the last one equals zero, of course).

Stage III. Searching for isomorphisms. Examining the activities of the type discussed above we next move on from relative frequencies to probabilities. Therefore we consider all the possible outcomes split into two parts, A and A':

A	(THH)	(HTH)	(HHT)	(TTT)
A'	(HTT)	(THT)	(TTH)	(HHH)

Since all the outcomes have the same probability we infer that $p(B)=\frac{3}{8}$, $p_A(B)=\frac{3}{4}$.

Now we can go back to the previous activities and use the same method to find conditional probabilities.

Stage IV. Representation. In stage III students calculate conditional probabilities simply by following the method of the first example (with three coins). Although the way of calculating conditional probabilities was always the same there was no general rule formulated. Now is the time to formulate the rule.

Let us assume that a set of all possible outcomes consists of N equiprobable outcomes. Let us consider now two events A and B and let us decompose the N outcomes according to the criterion of favourability to each of the events A and B. So we get our set of N outcomes sub-divided into four subsets. The cardinalities of the subsets are shown in the following table:

(*)

	B	B'
A	k	t
A'	m	n

$k+t+m+n=N$

The conditional probability is then equal to

(**)
$$p_A(B) = \frac{k}{k+t}$$

We may then ask the question - What is an adequate graphical representation of the conditional probability?

We may find it didactically attractive to replace the table (*) by a square diagram (Fig.12). It is a more general graphical representation because the outcomes need not be equiprobable.

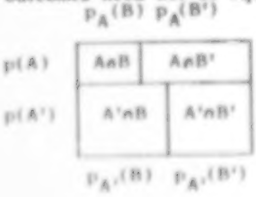


FIG.12

If the events A and B correspond to two successive stages of an experiment, a tree diagram is also a useful representation (Fig.13)

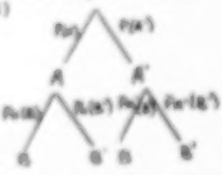


FIG.13

Stage V. The properties of the representation. In this stage we shall examine the representation in order to find properties of conditional probabilities ($P(A \cap B) = p(A) \cdot p_A(B)$ etc.). Studying the square diagrams one notices that in the case when $p_A(B)$ is independent of A, the diagram turns out as in Fig.14. Therefore square diagrams may be a good representation also of independent events. Didactically, there are two choices: either to teach conditional probability and independence together from the beginning (stages II-V) or to treat them separately in stages I-IV and to link them

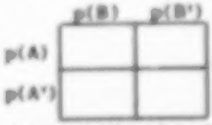


FIG.14

together in stage V.

EXPECTATION

Stage I. Initial interaction. Games and experiments considered now will always have the outcomes being rational numbers. Students will have an opportunity to gain some experience observing the duration of the game, the value of the win etc.

Stage II. Discovering regularities. We analyse each game separately and calculate experimental mean values in various situations. This can be motivated by asking questions about the fairness of the game etc.

Stage III. Searching for isomorphisms. If we compare the procedures used to calculate the mean values for the above experiments we will notice that the procedure follows the same pattern:

(*) $f_1w_1 + f_2w_2 + \dots + f_nw_n$ where f_i is the relative frequency of the outcome w_i.

Stage IV. Representation. Take the following problem: We want to play a game a large number of times. Can we predict or estimate the mean value before playing?

If the number of trials increases, the relative frequencies will come closer and closer to the probabilities. Thus we may think of the "ideal mean" which will replace the formula (*):

(**) $w_1p_1 + w_2p_2 + \dots + w_np_n$ where p_i is the probability of the outcome w_i.

Now we are ready to compute expectations in the above games and experiments and compare them with our experimental results.

Stage V. Properties of the representation. We will now examine the formula (**) and look for properties of the expectation. In other words we will now investigate all the above situations at the same time by simply examining their common representation.

The approach to the expectation proposed above, based on abstracting from concrete situations, has the advantage of avoiding the concept of a random variable which is difficult for beginners. Consequently, it makes it possible to talk about expectation much earlier, presumably as soon as students are able to assign the probabilities.

Later on, in more advanced courses, the above approach will be just a particular case where the outcomes are numbers and the random variable is an identity.

The special case considered here includes most of the typical examples of expectations in an elementary probability course. If we need the expectation of two random variables on the same space we simply think of two different experiments with different numbers as outcomes. This is sufficient at such an elementary level.

* * *

In a similar way, developing other important concepts (Bernoulli scheme, geometrical probabilities etc.) can be also put in the same framework of stages I-V.

As has been said in the introduction, a didactical consideration of isolated concepts is not enough. The problem now is to link them. This can, of course, be done in various ways. The optimum way seems to depend on the reasonable compromise between the following two postulates:

1. To start with the important concepts as soon as possible in order to provide sufficient time for deeper understanding.
2. For each concept stages I-IV should be in close succession to each other so that the student will not have forgotten the stage immediately preceding.

So, it seems to me that the compromise would be to have stage II immediately and stages III-V staggered, provided in the meantime the students can apply the knowledge they gain. The following table represents a possible order of concepts and stages. However, this has not yet been a subject of global experimental verification.

Concepts	Order of stages																		
Probability	I	II	III			IV		V											
Expectation				I	II											III IV			V
Multistage exp.					I		II		III		IV	V							
Condit. prob.					I		II			III			IV		V				
Independency							I			II			III	IV	V				
Bernoulli's sch.																	I-III	IV-V		
Random walks																	I-III		IV-V	

References

[1] Dienes, Z.P. (1971) An Example of the Passage from Concrete to the Manipulation of Formal Systems. Educ.Stud.in Math., 3.

[2] Engel, A. (1973) An Introduction to Probability. St.Louis, CEMREL.

[3] Engel, A. and Rade, L. (1973) Topics in Prob. & Stat., St.Louis, CEMREL.

[4] Engel, A. (1975) The Probability Abacus. Educ.Stud. in Math., 6.

[5] Engel, A., Varga, T. and Walser, W. (1973) Zufall oder Strategie.

[6] Fischbein, E. (1975) The Intuitive Sources of Probabilistic Thinking. Dordrecht, Holland, Reidel Publ. Company.

[7] Freudenthal, H. (1974) The Crux of Course Design in Probability. Educ.Stud. in Math., 5.

[8] Look on Luck. (1976) IOWO, Utrecht.

[9] Symposium on Combinatorics and Probability in Primary Schools (1975) (proceedings), Warsaw.

[10] Teaching of Probability and Statistics. (1970) L. Rade (Editor), Stockholm.

[11] Teaching Statistics and Probability. (1981) A.P. Shulte and J.R. Smart (Editors), National Council of Teachers of Mathematics, Yearbook.

DOES THE TEACHING OF PROBABILITY IMPROVE PROBABILISTIC INTUITIONS?

E Fischbein and A Gazit, Tel Aviv University

In the introduction to his well known treatise on probability, W. Feller has emphasized the necessity to develop adequate probabilistic intuitions as a basic component of the teaching of probabilities (Feller, 1968). But so far, little is known about probabilistic intuitions and their development under the influence of systematic instruction. In their main work on the chance concept in children, Piaget and Inhelder have concluded that, during the intuitive period (before the age of 6-7), the child is not able to distinguish clearly between chance and necessary phenomena. That distinction appears during the concrete-operational period together with elementary forms of probabilistic estimations. The concept of probability, as a formal construct, develops only during the formal operational stage and it represents a synthesis between the necessary and the possible (Piaget and Inhelder, 1951). This interpretation has been contested by various authors. Some of them consider that elementary probabilistic estimations may be found even in preschool children (Yost et al, 1962; Goldberg, 1966; etc.). Fischbein has also claimed that correct probabilistic intuitions may be identified in preschool children, but, because of social influences and present school curricula, only the deterministic aspect is actually developed (Fischbein, 1975). This situation may explain, partially, the fact that, frequently, the spontaneous probabilistic judgements of people are biased. Such incorrect evaluations of uncertain situations are often expressed as preconceived opinions and superstition. A good account of various classes of probabilistic misconceptions may be found in the work of Twersky and Kahneman (Kahneman and Twersky, 1972; Twersky and Kahneman, 1973). A more recent work dealing with children's probabilistic subjective evaluations is that of Falk et al. (1980).

The present investigation is concerned especially with the influence that the teaching of probabilities may have indirectly on intuitive probabilistic judgements. As mentioned above, there is very little information available about this problem. In an earlier work, Ojemann et al. have reported a positive, indirect effect of probability lessons on predictions made by their subjects (8-10 years old) in probability learning tasks (Ojemann et al., 1965).

The term "intuition" used in the present paper means, basically, a global synthetic, non-explicitly justified evaluation or prediction. Such a global cognition is felt by the subject as being self-evident, self-consistent, hardly questionable. For instance, the statement: "Through a point outside a line one and only one parallel may be drawn to that line". That statement is accepted intuitively, according to the above definition. An intuition is somehow a cognitive belief; sometimes, such cognitive beliefs may coincide with scientific accepted statements but sometimes they may not or may contradict them.

In the teaching process such intuitive attitudes must not be ignored. If they are correct they help the learner to acquire and integrate the corresponding scientific concepts. If they are not objectively acceptable they must be eliminated and adequate intuitive representations must be developed instead. If the teaching programme does not care about possible intuitive biases they will continue to mislead the learner despite the conceptual structures he has been taught.

Our point of view is that new intuitive attitudes can be developed only through personal involvement of the learner in a practical activity. Intuitions (cognitive beliefs) cannot be modified by verbal explanations. Therefore, a teaching programme which intends to develop an improved and efficient intuitive background for probability concepts and strategies along with the corresponding formal knowledge, must provide the learner with frequent opportunities to live actively, even emotionally stochastic situations. In such situations, he will confront his plausible expectations with empirically obtained outcomes.

The present research was chiefly devoted to evaluating the indirect effect of a course on probabilities on typical probabilistic, intuitively based, misconceptions.

The teaching programme itself, as a whole, was intended to determine the capacity of children 10-13 years old, to assimilate, correctly and productively, basic concepts and solving procedures in probability. According to our point of view the children have been presented with various situations which offered them the opportunity to be active in calculating probabilities, predicting outcomes, using operations with dice, coins and marbles for watching, recording and summing up different sets of outcomes. It is that reciprocal dynamics of theoretically computed probabilities and observed relative frequencies that may best contribute to the development of efficient probabilistic intuitions.

METHOD

The teaching programme (12 lessons) included: the concepts of certain, possible and impossible events, outcomes and events in a chance experiment: the concepts of chance and of measuring chances, the concepts of probability and relative frequency and the relation between them; counting outcomes; simple and compound events and their probabilities.

The subjects: a) the programme has been attended by 285 pupils; 70 grade 5 (2 classes); 160 grade 7 (5 classes) and 55 grade 7 (2 classes). Parallel control classes have also been investigated: 50 pupils (2 classes) grade 5; 200 pupils grade 6 (5 classes) and 55 pupils (2 classes) grade 7; (305 subjects in total in the control classes).

The assessment of the teaching effects has been obtained by means of two questionnaires: Questionnaire A was devised to assess to what extent the pupils, in the experimental classes, have assimilated and were able to use the concepts and procedures they have been taught. It has been administered only to the experimental classes. Questionnaire B was devised to assess the indirect effect of instruction on the children's intuitively based misconceptions. That questionnaire did not require special knowledge in probabilities and was administered to both experimental and control classes.

THE QUESTIONNAIRES

Questionnaire A (which will be described only briefly in the present paper) contained problems referring to all the notions mentioned above (the experimental lessons) such as: giving examples of certain, possible and impossible events; the probability of extracting a marble of a certain colour, when knowing the number of marbles of different colours contained in a box; the probability of various, simple and compound, events when rolling one or two dice; to give examples of simple and compound events in stochastic experiments; to sovle a problem based on proportional reasoning.

Questionnaire B, directly related to the topic of our paper, is presented below:

1. In a box there are 4 red, 3 green and 2 white marbles. How many marbles has one to draw in order to be sure that one will get at least one marble of each colour?

2. Joseph endeavours to enter the classroom each day by putting his right foot first. He claims that this increases his chance of getting good marks. What is your opinion?

3. Shula filled up once a lottery form with the following numbers: 1; 7; 13; 21; 22; 36, and she won. Therefore, she claimed that she must always play the same group of numbers, because it turned out to be a lucky one. What is your opinion on this?

4. Ruth prefers, when she participates in a lottery, to choose consecutive numbers like 1, 2, 3, 4, 5, 6. She claims that in this way she increases her chance of winning. On the other hand Jenny claims that the chance of getting six consecutive numbers like 1, 2, 3, 4, 5, 6 is smaller than that

of getting a random sequence of numbers. She says that a lottery is something chancy and therefore there is no chance of getting a sequence of consecutive numbers. What is your opinion with regard to the two attitudes, that of Ruth and that of Jenny?

5. Gill has participated in a weekly lottery during the last two months. So far he has never won but he has decided to go on for the following reason: "Lottery is a game based on chance, sometimes you win sometimes you lose. I have already played many times and I have never won. Therefore, I am sure more than before that I shall win in one of the next games". What is your opinion with regard to Gill's explanation?

6. Inbar has 100 white marbles and 50 black ones in a box. Itay has in his box 200 white marbles and 100 black ones. Without looking, each of them takes a marble from his own box. Compare their chances of extracting a black marble. Mark with an x the correct answer:

A. Inbar has a bigger chance of extracting a black marble.
B. Itay has a bigger chance of extracting a black marble.
C. Their chances of extracting a black marble are equal.
D. It is impossible to get an answer from the above data.

7. Gilla is 10 years old. In her box, there are 40 white marbles and 20 black ones. Ronit is 8 years old. In her box there are 30 white marbles and 15 black ones. Each of them draws one marble from her own box, without looking. Ronit claims that Gilla has a greater chance of extracting a white marble because she is the older one, and therefore she is the cleverest of both of them. What is your opinion on this?

8. Uri has, in his box, 10 white marbles and 20 black ones. Guy has in his box 30 white marbles and 60 black ones. They play a game of chance. The winner is the child who pulls out a white marble first. If both simultaneously take out a white marble no one is the winner and the game has to go on. Uri claims that the game is not fair because in Guy's box there are more white marbles than in his box. What is your opinion on this?

Question 1 checks the capacity of the subjects to identify an outcome which is certain in a stochastic experience. Questions 2 and 3 refer to the idea of luck, in fact to the acceptance of a prediction made in disregard to a fair evaluation of chances. Question 4 refers to a typical probabilistic misconception called by Twersky and Kahneman (1972, p.134) "representativeness". A sequence of consecutive numbers appears, intuitively, to be less frequent in chance extractions than a random one because the first seems to be less "representative". Question 5 deals with what is called a negative recency effect. The intuitive conception on which that effect is based is that of a necessary equilibration, in the long run, between the frequencies of various possible outcomes. It expresses a confusion between the probability of getting a single result (which, in fact, does not depend on the previous outcomes) and the probability of getting an expected set of outcomes which depends on the structure of the set (see Fischbein, 1975, pp.58-62). Question 7 checks a misconception described by Falk et al. 1980: children may assign a role, in chance events, to the personal qualities of the player, though, objectively such an effect does not exist. Questions 6 and 8 deal with proportional reasoning. Proportional reasoning is considered to be a basic tool of probabilistic reasoning. We wanted to check this relation.

RESULTS

We intend to describe in detail the data obtained with Questionnaire A in a different paper. Nevertheless, a brief presentation of the basic results obtained with A seems to be necessary (the conceptual gains as a direct result of instruction).

Questionnaire A. There is a steady increase with age of proportions of correct answers, with differences reaching as big as 30-40% between two consecutive grades. The majority of the children of grade 5 did not perform well at most of the items (except those requiring examples of certain and possible events). Items referring to computations of probabilities of simple and compound events were solved by 60-70% of the subjects in grade 6 and by 80-100% in grade 7. For instance with regard to the question: "What is the probability of obtaining "1" when rolling a die?" we obtained the following percentages of correct answers: 53.5 in grade 5; 87.5 in grade 6; 100 in grade 7. The probability of obtaining a number smaller than 6: 32.4 in grade 5; 63.7 in grade 6; 92.3 in grade 7. Computations referring to compound events expressed in sums were solved by 10-20% of the subjects in grade 6 and by 30-40% in grade 7 (Ex.: One rolls a pair of dice and one sums up the figures obtained. What is the probability of obtaining an even number? Percentages of correct answers: 2.8 in grade 5; 16.1 in grade 6; 44.2 in grade 7). To the question referring to proportions, the following percentages of correct answers have been obtained: 1.4 in grade 5; 23.2 in grade 6; 44.2 in grade 7.

Questionnaire B. Question 1 refers to the notion of certain events. The data appear in Table 1. This question does not refer explicitly to probabilistic estimations. But, as a matter of fact, it checks the capacity of the child to produce, mentally, a situation which corresponds to a certain event and this is one of the basic concepts of the theory of probabilities.

As one can see by inspecting Table 1, there is a steady improvement with age for both control and experimental classes. The experimental classes got, at each level, higher percentages of correct answers than the control ones.

TABLE 1

Percentages of correct and incorrect answers for question 1[*]

	Grades					
	5		6		7	
Answers	E	C	E	C	E	C
Correct: Eight marbles	14.1	2.0	31.5	21.8	53.8	45.5
Incorrect number of marbles:						
3	23.9	26.5	23.2	23.8	11.5	18.2
6	12.7	22.4	11.3	13.9	19.2	25.5
2 or 1	32.1	34.7	9.6	7.9	5.8	5.4
7	1.4	0.0	3.6	1.5	5.8	1.8
Total incorrect answers	70.4	83.6	47.7	47.1	43.3	50.9
No answer	15.5	14.4	20.8	31.1	2.9	3.6

[*] E stands for experimental and C for control classes.

Comparing the figures, it turns out that factor age is more effective than factor instruction in the given conditions (control classes grade 7 perform better than experimental classes grade 6).

Question 2: The subjects have been asked whether entering the classroom with the right foot will increase their chances of getting high marks in the respective day (see Table 2).

TABLE 2

Percentages of correct and incorrect answers, Question 2

	Grades					
	5		6		7	
Answers	E	C	E	C	E	C
"No" correct	60.0	81.4	89.9	72.3	97.3	83.0
"Yes" (incorrect)	9.0	11.8	3.6	22.8	1.9	13.8
No answer	31.0	6.8	6.5	4.9	0.8	3.2

Generally, this misconception has been, spontaneously, rejected by most of the children at each age level. In the experimental classes there is an increase with age of the proportions of correct answers, an important leap taking place from grade 5 to grade 6. It is clear that in grade 5 instruction did not produce any improvement in those children who first hesitated or accepted the misconception. On the contrary, much more non-instructed children rejected the superstitious claim than the instructed ones. According to this data it turns out that the teaching process - probably inadequate to that age - has had a negative effect on the children's chance estimations!

On the other hand, in grades 6 and 7 the experimental classes performed better than the control ones - although the differences were not very high. Nevertheless it is a very important result: it proves that the process of teaching probabilities may have not only a formal-instructive effect (transmission of knowledge) but also an indirect positive effect on the children's intuitive estimations in uncertain situations.

Question 3 asked whether in a lottery game a certain group of numbers which once won may be considered as being an especially fortunate one (see Table 3).

TABLE 3

Percentages of correct and incorrect answers - Question 3

'The fortunate number'

	Grades					
	5		6		7	
Answers	E	C	E	C	E	C
"No" (correct)	21.1	12.2	52.4	40.6	53.9	52.7
"Yes" (incorrect)	0.0	4.1	2.4	6.9	0.0	7.3
The same number cannot win any more (incorrect)	47.9	63.3	29.2	46.0	21.2	29.1
No answer	31.0	20.4	16.0	6.5	24.9	10.9

A highly frequent mistake consists of considering that a number which once won will not win any more. In both control and experimetal classes there is a clear drop with age of this type of basic misconception. Age seems then to be an essential factor. Very few subjects admit the possibility of a lucky number. At each level, experimental classes perform better (more correct answers) than control ones.

Question 4 asked about the chances that a group of consecutive numbers will win in a lottery game. See Table 4 for results.

TABLE 4

Percentages of answers referring to the probability of a set of consecutive numbers winning in a chance game

	Grades					
	5		6		7	
Answers	E	C	E	C	E	C
There is no relation (correct)	21.2	38.8	27.4	24.2	32.7	12.8
Random numbers have a higher probability of winning (incorrect)	28.2	34.7	44.0	52.5	42.3	41.8
Lottery is a chancy game (ambiguous)	5.6	10.2	16.1	10.4	25.0	40.0
No answer	45.0	16.3	12.5	12.9	0.0	5.4

Many children do not refer directly to the probability that a group of consecutive numbers will win but give an ambiguous answer ('lottery is a chancy game') which may be interpreted in two opposite ways: (a) A lottery is a chancy game and, therefore, the nature of the numbers does not influence the probability of winning or not; (b) A lottery being a chancy game, a group of consecutive numbers has a smaller chance of being extracted.

The proportion of clearly correct answers (the possibility of a set of numbers being extracted does not depend on its content) is higher in the experimental classes 6 and 7 than in the corresponding control classes. The difference is impressive in grade 7. One can also see that in the experimental classes the proportions of correct answers grow steadily with age (grade) while in the control classes there is a decrease with age in the proportions of subjects who answer correctly. The presence of the ambiguous type of answer ('A lottery game is chancy') makes a further interpretation of the results rather difficult. One may say that in general, with age, children become more aware of the non-predictability of chance phenomena (as opposed to certain ones). But, at the same time, the internal conflictive nature of the intuition of chance becomes more acute with age. Intuitively, a lottery game being a chancy one, every set of numbers must have the same probability of winning. At the same time, one can hardly accept, intuitively, that the set 1, 2, 3, 4, 5, 6 has the same chance of winning as an "irregular" group of numbers. The picture in grade 7 is highly instructive. In the control classes about 40% of the subjects prefer non-consecutive numbers, about 40% affirm that a lottery is a chancy game (ambiguous) and only 12.8% admit, explicitly, the independence of chance in relation to the nature of the numbers. In the experimental classes, again, about 40% prefer a random set of numbers. But this time only 25% answer ambiguously while 32.7% affirm explicitly the independence of probability. This is a drastic change of a very complex intuitive attitude, as an indirect effect of an instructive process.

With regard to the categories of answers which may be characterised as ambiguous ones, one can see that there is a drop at grade 7 especially in experimental classes.

Question 5 was intended to check the childrens' attitude with regard to the same problem as question 4: the independence of probabilities. This time, the subjects had to refer to the independence of probabilities in a sequence of trials which, objectively, leave the probabilities unchanged. (Gill has not won so far, though he has tried many times. He is convinced that he must win in one of the following games). See Table 5 for results.

TABLE 5

Percentages of various categories of answers to question 5: previous successive losses may increase the probability of a further positive outcome?

	Grades					
	5		6		7	
Answers	E	C	E	C	E	C
"No" (correct)	18.3	18.4	44.6	33.7	73.1	49.0
"Yes" (error)	12.7	22.4	10.7	21.3	5.8	14.5
"Lottery is chancy"	18.3	28.5	23.8	23.2	9.6	14.5
Undecided ("possible")	9.9	12.2	14.9	16.8	7.7	18.2
No answer	40.8	18.5	6.0	5.0	3.8	3.8

There is a clear and steady increase with age (grade) in proportions of correct answers ("The odds are the same at every trial") in both experimental and control classes. On the other hand, in grades 6 and 7 the proportions of correct answers in the experimental classes are neatly higher than in the control classes. In grade 5 they are equal. It is, therefore, clear that both factors, age and instruction, are acting in the same direction, leading to a more adequate intuitive approach with regard to chance estimations in sequences of trials. We must again emphasise that instruction acted here only indirectly: no direct mention has been made during the experimental lesson of that matter.

Question 6 deals with the comparison of probabilities based on a comparison of ratios: 100 white and 50 black marbles versus 200 white and 100 black marbles. The subject has to compare the odds to extract a black marble. (See Table 6).

TABLE 6

Percentages of correct answers when comparing the odds to extract a black marble from (100 w; 50 b) versus (200 w; 100 b)

	Grades					
	5		6		7	
Answers	E	C	E	C	E	C
The odds are equal (correct)	21.1	22.4	42.9	51.5	80.8	85.5
In 200/100 the odds are bigger (error)	32.4	20.4	36.9	20.3	15.4	9.1
In 100/50 the odds are bigger (error)	8.5	6.1	6.5	4.0	1.9	0.0
One cannot decide (error)	11.3	12.9	6.5	21.8	0.0	3.6
No answer	26.3	8.2	7.2	2.4	1.9	1.8

There is an increase with age (grade) in proportions of correct answers in both experimental and control classes almost doubling themselves from grade to grade. The age factor acts very strongly in this respect. But, at the same time, at all three levels considered, the control classes performed better than the experimental ones: Instruction had rather a negative effect. Though the differences are not very high, this tendency seems to be a systematic one.

Specifically, there are higher proportions of children who choose the box with 200 W/100 B as favouring the extraction of a black marble. This means that for many chidlren (20-30%) the absolute quantities are still decisive with regard to the probability of chance events.

Question 7. Do intelligence and age influence the outcome of a chance event? See Table 7 for results.

TABLE 7

Percentages of various categories of answers - "Do intelligence and age influence the outcome of a chance event?"

	Grades					
	5		6		7	
Answers	E	C	E	C	E	C
"No" (correct)	49.4	63.2	67.8	84.2	82.6	87.3
"Yes" (error)	14.1	10.2	22.6	9.9	9.6	3.6
No answer	6.5	26.6	9.6	5.9	7.8	9.1

The majority of subjects, at each level, reject the claim that intelligence and age may affect the outcome of a purely chancy game (except the experimental classes of grade 5). There is an increase in the proportion of correct answers with age in both experimental and control classes. But, very surprisingly, there were higher percentages of correct answers in the control classes. Again, a certain factor (objectively a non-relevant one) is considered by a number of children as being effective in deciding the outcome in a chance situation. Instruction seems to have played rather a negative role with regard to that problem.

The last question (the 8th) was devoted again to a comparison of odds based on a comparison of ratios (10/20 versus 30/60) but using a different wording from that used in question 6 (see Table 8). Does the fact that in (30W/60B) there are more white marbles than in (10W/20B) determine a higher probability of extracting a white marble from the first box?

TABLE 8

Categories of answers in percentages. Comparison between the chances of extracting a white marble from (10W/20B) versus (30W/60B)

	Grades					
	5		6		7	
Answers	E	C	E	C	E	C
The chances are equal (correct)	21.1	28.6	53.0	50.4	69.2	72.7
The equality of ratios made explicit*	(11.3)	(10.2)	(38.7)	(26.7)	(65.4)	(61.8)
Higher chances for 30W/60B (incorrect)	35.2	40.8	35.7	35.1	17.3	12.7
No answer	43.7	30.6	11.3	14.5	13.5	14.6

* The figures in paranthesis are extracted from those referring to the previous category (the chances are equal) in which they are also included.

It seems that the teaching of probability according to our programme did not improve the capacity of the pupils to make use of proportional reasoning for reaching a correct estimation of chances. Again, the picture presented by grade 7 is highly instructive. There are a few more subjects, at that level, in the experimental classes, who are able to justify their correct answer by referring explicitly to the equivalence of ratios. But, despite this, there are also more subjects in the control classes who answer correctly. The differences are very small, statistically non-significant. It is clear that the teaching programme did not improve the global estimation of odds when proportionality was involved. But some of those pupils who were ready to estimate correctly the equality of chances have learned, as a result of our programme, to use, explicitly, the formal justification (equality of ratios). Considering only those subjects who justified explicitly their correct answers one can see that, at each age level, there were more subjects in the experimental classes than in the control classes who were able to offer such an explanation.

The factor age seems to play a decisive role: the increase with age in the proportions of correct answers is evident, the great leap taking place from grade 5 to grade 6.

SUMMARY AND CONCLUSIONS

1. There is a clear increase with age (grade) in proportions of correct answers in both experimental and control classes at all the items. While 5th grade children (10-11 years old) seem not to be intellectually ready to assimilate the basic concepts and procedures of probability reasoning, 6 graders seem to be mentally ready for that task. About 60-70% in grade 6 and about 80-100% in grade 7, performed well at almost all the items referring to elementary questions with simple and compound events.

2. With regard to the indirect effect of teaching probabilities on the intuitive estimations of children in uncertain situations the picture is rather complex: a) There were two typical biased intuitions on which the teaching process has had a clear improving effect (in grades 6 and 7): the 'representativeness' tendency, described by Tversky and Kahneman and the well known negative recency effect. As a matter of fact, there were still about 40% of subjects in grades 6 and 7 who believed that a set of consecutive numbers had a smaller chance of winning than an 'irregular' set of numbers. But, at the same time, there were more subjects who answered correctly in the experimental classes than in the control ones.

A third question with regard to which the teaching process had an improving effect was that referring to a superstition: would the fact of entering the class first with the right foot ensure higher marks? That claim is rejected by most of the pupils, but in grades 6 and 7 more subjects from the experimental classes than from the control ones rejected it.

There was one item with regard to which the control classes performed clearly better than the experimental ones. Do age and intelligence increase the chances of winning in a (fair) chancy game? The majority of subjects rejected the possibility of such an influence. Nevertheless it was surprising to see that while 84.2% of subjects answered correctly in the control classes, in the experimental classes only 67.8 answered correctly (grade 6).

A possible explanation may be the following: There is an intuitive correct estimation in children (aged 12-13) of the fairness of a chancy game (in this case based on proportions). Instruction may have drawn the attention of the subjects to the fact that even in chancy games, some kind of prediction is possible based on reasoning and knowledge. Perhaps some of our subjects have been confused by this intervention. This phenomenon deserves further research.

3. Two items were devoted to proportional reasoning. In relation to both items, the experimental classes performed, generally, worse than the control ones. Questionnaire A referred also to a proportion problem and even in grade 7, 60% of the subjects were not able to answer it correctly. These findings are in accordance with those frequently mentioned in recent years in relation to the same topic. (See Karplus, 1976; Hart, 1981, pp.66-101). In our opinion, an important theoretical conclusion may be drawn from the above data: Probabilistic thinking and proportional reasoning are based on two distinct mental schemata. A progress obtained, as an effect of instruction, in one direction does not imply an improvement on the other. Certainly, probability computations may require ratio comparisons and calculations, but probability, as a specific mental attitude, does not, necessarily imply a formal understanding of proportion concepts. This finding is in accordance with other results obtained before. Yost et al. (1962), Goldberg (1966) have shown that one can identify, even in preschool children, correct probability estimations while proportional reasoning is still deficient. But it is also possible that at a very basic intuitive level, the two schemata share the same roots, namely an intuition called by us the intuition of relative frequency, which is revealed especially in probability learning tasks (Fischbein, 1975, pp.58-65). It refers to the subject's capacity to adapt intuitively his reactions to the relative frequencies of a group of stimuli. Such a common base may explain

Ojemann's finding quoted above. But, in fact, the two forms of reasoning, the probabilistic and the proportional ones, include, besides this hypothetical common root, clusters of intuitions and technical procedures which make them distinct. By emphasizing (via systematic instruction) specific probability viewpoints and procedures one may disturb the subject's proportional reasoning, still fragile in many adolescents. When we implemented our program we were not yet aware about that danger. However, we suppose now that, by using adequate means, this negative side effect can be avoided.

References

[1] Falk, R., Falk, R. and Levin, J. (1980) A potential for learning probability in young children. *Educational Studies in Mathematics*, 11, 181-204.

[2] Feller, W. (1968) *An introduction to probability theory and its applications.* John Wiley, New York.

[3] Fischbein, E. (1975) *The intuitive sources of probability thinking in children.* D. Reidel, Dordrecht.

[4] Goldberg, E. (1966) Probability judgement by pre-school children; task, condition and performance. *Child Development*, 37, 157-167.

[5] Hart, K.M. (Editor) (1981) *Children's understanding of mathematics.* John Murray, London.

[6] Kahnemann, D. and Tversky, A. (1972) Subjective probability; a judgement of representativeness. *Cognitive Psychology*, 3, No.3, 430-454.

[7] Karplus, R. and Karplus, E. (1971) *Proportional reasoning and control of variables.* Cambridge, Mass.: Division for Study and Research in Education.

[8] Ojemann, R.H., Maxey, E.J. and Snider, B.C. (1965) Effects of guided learning experiences in developing probability concepts at the fifth grade level. *Perceptual and Motor Skills*, 21, 415-427.

[9] Piaget, J. and Inhelder, B. (1951) *La genèse de l'idée de hasard chez l'enfant.* P.U.F. Paris.

[10] Tversky, A. and Kahneman, D. (1973) Availability: a heuristic for judging frequency and probability. *Cognitive Psychology*, No.5, 207-232.

[11] Yost, P., Siegel, A.E. and Andrews, J.M. (1962) Non-verbal probability judgement by young children. *Child Development*, 33, 769-780.

APPENDIX

Chi square tests for comparisons of frequencies of correct answers between experimental and control classes separately for each grade level (df=1)! and each item of Questionnaire B.

Grades

	5		6		7	
Questions	Chi square	P	Chi square	P	Chi square	P
1	3.91	$P<0.05$	3.92	$P<0.05$	N.S.	
2	5.63	$P<0.02$	24.6	$P<0.001$	5.39	$P<0.05$
3	N.S.		4.19	$P<0.05$	N.S.	
4	N.S.		N.S.		5.20	$P<0.05$
5	N.S.		4.03	$P<0.05$	5.53	$P<0.02$
6	N.S.		N.S.		N.S.	
7	N.S.		19.80	$P<0.001$	N.S.	
8	N.S.		N.S.		N.S.	

STOCHASTICS TEACHING BASED ON COMMON SENSE

H-J Bentz, University of Osnabrück

The following list could give a rough description of contents and goals of usually practised stochastics courses (for simplicity, this list will later also serve as a guide-line for this presentation):

- Discussion of the properties of P (Probability), comparison with f (relative frequency
- Algebra of events (analysis and composition of events)
- Laplace-P, Combinatorics, Bayes'-Formula, Independence
- Random variables (RV), Expectation (E), Variance (Var)
- Distributions, e.g. Binomial, Hypergeometric, Poisson, Gauss or Normal Distribution
- Chebyshev's Inequality, laws of Large Numbers
- Descriptive Statistics
- Tests (Parameters, Independence, Chi-Square, etc.)

Normally the theoretical background is described by Kolmogoroff's Axioms, the Algebra of sets and Measure theory.

We, the teachers, know about difficulties (in learning, understanding and handling) which may arise at nearly every stage during the course. The reasons are various. Partly they come from misconceptions (or no conceptions), misfeelings, lacks in techniques, difficulties in understanding and transforming literal texts, even from psychological obstacles - to name only some of those deficiencies which I observe again and again.

Maybe some of the very important contents of our courses could be taught with more success (concerning the students) and less effort (concerning the teacher) by using a different concept. The following outline (which is based on a joint paper [1] with G. Palm) will describe a stochastics course which essentially differs from the usual ones, namely e.g.:

in theory:	it does not use the sample space Ω, the algebra of events, set- and measure-theory, nor Kolmogoroff-Axioms, but is based on the concept of a random-variable and its expected value.
in concepts:	Probability is not described by axioms or by comparison with frequencies, but is defined as (is the name of) a certain expected value. So, expectation is not defined as $\Sigma\, p_i\omega_i$, but is one of the fundamental notions itself.

in proceeding: the aim is to "analyse" random variables (RV), i.e. to simplify RV by decomposition in simpler parts. So, composition and decomposition play an important role.

in notation: a certain notation which allows us to describe decompositions is introduced. In addition it is very helpful in transforming literal texts into "mathematical" symbols.

in models: coins, dice, cards, urns, ..., are not used as standard models, but a special "wheel of fortune" serves as first and global model. this "spinner" has a front on which one can write with chalk and wipe clean again. With it (practically) all RV can be represented, even continuous ones.

The reader should not feel deterred by this list of (as I mean partly severe) differences; the following explanations will take care of the contents mentioned at the very beginning. But the key-words in the more detailed description below will be: Random Variable (RV), Expectation (E), Decomposition, Simple and, last but not least, Spinner (SP).

The text is divided in places into global remarks on the theoretical background and the procedure in the classroom. So it is rather more a description than an instruction for the students, although many examples and hints mentioned stem from classroom work.

I. Basic notions: RV, W and E. The space *RV*

A RV is to be looked at as an accidental outcome (= numbers) of an (actual) experiment. The E of complicated RVs has to be determined by use of simpler RVs. So the aim will be to analyse RVs and to decompose them. What do "decomposition" and "simple" mean? The space *RV* of all RVs, what structure should it have?

Each RV X has outcomes. These numbers are collected in a set W(X). Besides this, X has an expected value. So there have to be two mappings, a positive linear functional $E : \mathit{RV} \to \mathrm{IR}$,
and an unspecified $W : \mathit{RV} \to$ Power set (IR).

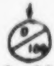
Simple then means $W(X) \leq \{0,1\}$, *decomposition* means linear combination. Therefore *RV* should be an ordered vector space at first (and a Jordan-Banach-Algebra later). More details are given by axioms, see [1].

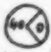
It should already be stressed here that the course continually utilises the property $E(X+Y) = E(X) + E(Y)$, i.e. the linearity of E.

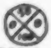
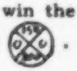
As a model for RVs a spinner is used. With it the main properties of RVs and important stochastic principles are discussed. RVs upon spinners are also called *games*.

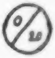
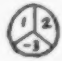
II. Learning and playing of games

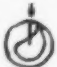
First examples:

1. Introduce games like . The spinner is divided into two equal parts, the possible outcomes are 0, 100, so W(•) = {0,100}. The questions are, what will we expect to win on the average? What amount would we ask for, to offer this game to a player. What amount would we bet on this game? ... Continue with similar examples.

2. Introduce the *sum* of games, e.g. + . One will win the sum of the (not necessarily independent) outcomes. W = {0, 40, 60, 100} if both run independently of each other. W may vary according to the type of dependence.

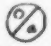
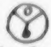
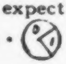
3. Introduce the *multiple* of a RV, e.g. 3. . One will win the outcome of the spinner times the factor. Note, this RV is the same as . W = {0, 150, 300}.

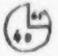
4. Introduce the *product* of RVs, e.g.  • . W = {0, 20, 40, -60} if both run independently.

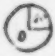
5. (Eventually) introduce games of continuous type, e.g. .W = [0,100], here one wins that fraction of the radius which is "cut" by the curve (total length ≙ 100). In this respect this model is in fact different from urns, dice, coins, ...

With these examples and types the students get a first impression about RVs. The main emphasis in this stage lies on the "learning" of RVs. The meaning and properties of E will be examined afterwards.

III. Expectation

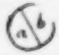
Start with games like , , ..., i.e. games with two partitions and W = {0,a}. Here, E is easily found (≙ average outcome). Students either use symmetry or statistical arguments. "Half of the time we get a, half 0, so we expect a/2". After a while, E is immediately found, if the game is of type a • .

Continue with the problem X = , E(X) = ?

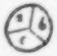
1st solution: $X = 40 +$ , the latter is well known, so $E = 40 + 5$.

2nd solution: "statistical argument" : $40 \cdot 3/4 + 60 \cdot 1/4$.

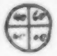
3rd solution: $X =$, both run strictly synchronous. $E = E_1 + E_2$.

4th solution: "à la *Huygens* ". This method is to be explained by a simpler game, namely $Y =$ 

Two partners are assumed to play against each other, both stake x. Let it be agreed that the one who wins shall give a to the one who loses. This agreement is fair. Everyone has the same chance to get a (loser) or to get $2x-a$ (winner). To have the game Y, set $b = 2x-a$, then one has the same chance for a and b.

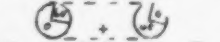
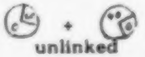
As consequence $x = (a+b)/2$ for the value of the game Y.
In like manner one obtains $x = (a+b+c)/3$ as the value of the game 

(Three partners have to stake x with the agreement to pay b to the second and c to the third, if the first is winner, etc.).

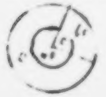
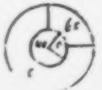
With these preliminaries our X from above equals 
and $E = (40 + 40 + 40 + 60)/4$.

(For a detailed description and various remarks, see [3], [6]).

Remark: It should be stressed that E remains invariant (as long as the total sum of the outcomes remains unchanged and the partitions are all equal), following Huygens' arguments, even if the single amounts or parts of them are distributed arbitrarily in the spinner-field.

5th solution:  $\overset{\text{in E}}{\underset{?}{=}}$ , right side is familiar.

The equality sign does not hold for the games themselves, as W("left") = {40,60} ≠ W ("right") = {0,40,60,100} but it holds in E. This very fundamental fact should be discussed extendedly. Hints for procedure: Look at the left and there at X_1. Its E_1 is already known, $E_1 = 15$. This E_1 is not influenced by the other game, because the outcomes are determined only by chance. The analogue holds for the second game on the left, its E_2 equals 30. So, if both outcomes were to flow into one single "money-box", $E_1 + E_2$ would be expected to be in there. This argument also shows that the total E does not depend on the initial position of the two separate games. Altogether one gets the following understanding: If two games are played independently (with two spinners) the total E is the same as if they are played with one spinner, e.g.

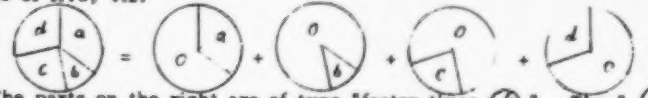
 , 

so the rule $\boxed{E(X+Y) = E(X) + E(Y)}$ holds in general. Check your intuition by the question: "How would I like to link the two games X_1, X_2 ?". (Note that there is a certain correspondence between this "equality in E" of the examples in the brackets, just above, and those stated as remark to Huygens' solution).

Some didactical comments to this 5th solution are given in [2].

IV. Decomposition

All the above solutions are worth mentioning in the classroom. We prefer to continue the course, referring to the last one. The advantage lies in the generality of its application. We use it in a straightforward manner to determine Es of RVs, viz.



, and so on.

The parts on the right are of type "factor times ". The " " is already familiar to us and is now called *simple* RV.

Its W is a subset of {0,1}, so e.g. ①, ⓪, are simple games.

V. Notation

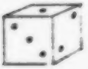
Simple RVs are written in the form $1_{[A]}$. The subscript "A" is an event or a condition (given in everyday language or in mathematical symbols). If A turns out to be true, then $1_{[A]}$ assumes the value 1, and it will be zero in the other case ("characteristic function").

Notes: (a) $1^2_{[A]} = 1_{[A]}$, so $1_{[A]}$ is idempotent. (b) 1 := ①.

Examples. $1_{[\text{dropping a coin, heads will be uppermost}]} =$

$1_{[\text{the sum of two dice will be 1}]} = 1_{[2>7]} =$ ⓪

$1_{[\text{rolling a die, we will see a number less than 10}]} =$ ① .



$= D = 1_{[\text{1 comes up}]} + \dots + 6 \cdot 1_{[\text{6 comes up}]}$

$= 1_{[D=1]} + 2 \cdot 1_{[D=2]} + \dots + 6 \cdot 1_{[D=6]}$ (*"standard dec."*, see below)

$= 1_{[D \geq 1]} + 1_{[D \geq 2]} + \dots + 1_{[D \geq 6]}$

As the die example shows the decomposition is not unique in general. We draw attention to one specific form. Let X be a RV with $W = \{a_1, a_2, \dots, a_n\}$. Then the decomposition $X = a_1 \cdot 1_{[x=a_1]} + a_2 \cdot 1_{[x=a_2]} + \dots + a_n \cdot 1_{[x=a_n]}$ is called the *standard decomposition* of X.

It is a central (and sometimes "funny") point in this concept to find the best (adequate, economic, ...) decomposition, according to the information given in the problem. Therefore, decompositions (and compositions) should be taught in the classroom. Some examples will serve as demonstration.

Example 1. A die is rolled until "3" comes up. Average waiting time ? T is the number of trials which have to be conducted until 3 comes up for the first time. This is a RV with the standard decomposition

$T = 1_{[T=1]} + 2\cdot 1_{[T=2]} + \ldots + n\cdot 1_{[T=n]} + \ldots$. We find

$E(T) = \frac{1}{6} + 2\cdot\frac{5}{6}\cdot\frac{1}{6} + \ldots + n\cdot(\frac{5}{6})^{(n-1)}\cdot\frac{1}{6} + \ldots$, which is not immediately calculable by our students.

Compare this with the decomposition

$T = 1_{[T\geq 1]} + 1_{[T\geq 2]} + \ldots + 1_{[T\geq n]} + \ldots$, and

$E(T) = 1 + \frac{5}{6} + \ldots + (\frac{5}{6})^{n-1} + \ldots = 6$ immediately.

Example 2. (see [5]). A bag contains the four numbers 1, 2, 3, 4 which are drawn in random order, every number in the correct place winning £1. Compute the expectation of this promise.

The RV is called F. Its standard dec. is given by

$F = 4\cdot 1_{[1234]} + 2\cdot 1_{[1243 \text{ or } 1324 \text{ or} \ldots \text{or } 4231]} + 1_{[1342 \text{ or} \ldots \text{or } 2314]}$

$E(F) = 4\cdot\frac{1}{24} + 2\cdot\frac{6}{24} + \frac{8}{24} = 1.$

Compare this with

$F = 1_{[1 \text{ at the first draw}]} + 1_{[2 \text{ at second}]} + \ldots + 1_{[4 \text{ at last draw}]}$

$E(F) = 1/4 + 1/4 + \ldots + 1/4 = 1.$

Of course, the standard decomposition is not always a "bad" one, but it is suggested in the "classical" courses because there we usually start with the sample space. In our course, decomposition is systematically trained, and there is no "favourite" generally given decomposition. (Look back again at Example 2, but now with 50 numbers, say. Working with the standard dec. soon turns out to be hopeless, so one is forced to look for a more economical one).

Remarks: The notation $1_{[\ ]}$ combined with the principle of (linear) decomposition seems to be very helpful in transforming literal texts into a form where the E-rule finds application. Remember that subscript-events are allowed to be formulated as well in ordinary language as in mathematical symbols. So this takes greatly into account the different abilities of the students. Very often it is possible to transform the problem directly in terms of simple RVs (see Example 2, second solution). This was one reason for the label "Common Sense" Probability.

VI. Logical rules

It may happen that the event or condition "A" itself is a composition of other, simpler ones. To handle such cases the following rules should be kept in mind:

(1) $\boxed{1_{[A]} = 1 - 1_{[\text{not } A]}}$ (2) $\boxed{1_{[A \text{ and } B]} = 1_{[A]} \cdot 1_{[B]}}$

(3) $\boxed{1_{[A \text{ or } B]} = 1_{[A]} + 1_{[B]} - 1_{[A \text{ and } B]}}$ (proofs by trivial computation).

These identities reflect certain rules, which are well known to students but in terms of set algebra. However, (1), (2), (3) differ from those in several respects, namely:

1.) They show that characteristic functions of events "behave" like numbers; that simple games carry arithmetical properties.

2.) At this stage, these rules are easily motivated.

3.) Their inherent decomposition should be utilised, and in fact, E applies immediately in rules (1) and (3).

4.) The question if (and under which conditions) E also applies to (2) - in general: when does $E(X \cdot Y) = E(X) \cdot E(Y)$? - directly leads to the concepts of independence and dependence.

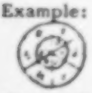
So, the algebra of events - often attacked for didactical reasons if posed as an introductory chapter to probability courses - may turn out to be favourable if used as "logical rules" in a context like this.

VII. Probability

In stochastic problems the word 'probability' is frequently used. We should not miss it. In this course *probability of an event* is defined in the following manner: $P(A) := E(1_{[A]})$.

So, the simple RVs provide a connection with 'probability'. P of an event A means the expected value of the attached simple game $1_{[A]}$.

Example: The probability of 5, rolling a die, is found as follows:



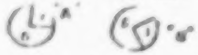
The expected value of the "inner" game is 1/6, so $P(5) = 1/6$.

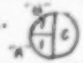
(Note: Using the spinner, the distinction between "probability of an event" and "value of the outcome" is obvious).

VIII. Stochastic Independence

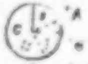
Students tend to apply the E-rule to $1_{[A \text{ and } B]} = 1_{[A]} \cdot 1_{[B]}$ as well, getting $E(1_{[A]} \cdot 1_{[B]}) = E(1_{[A]}) \cdot E(1_{[B]})$.

At this point a discussion about "independence" should start. Here we use two spinners. The game $1_{[A \text{ and } B]}$ played on two unlinked spinners, e.g.



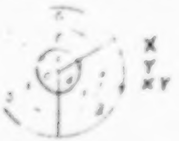
has an E which differs (in general) from the E of $1_{[A \text{ and } B]}$, played on a single spinner, e.g. 

There is no doubt about this, because the information "A has come out to be true" does not influence our betting on "B" in the first case, whereas the same information changes our betting in the second case.

Note: The "overlap" of "A" and "B" in the second case may be chosen in such a manner that both E's are equal, and then the information on "A" does not tell anything about "B", just as in the unlinked situation, viz. 

This leads to the definition: Two simple RV_s X, Y are *independent* if $E(X \cdot Y) = E(X) \cdot E(Y)$ holds. Otherwise they are *dependent* (on each other). In general - if X, Y are not simple - the definition uses their standard-dec. (in the discrete case): $X = \sum_i a_i \cdot 1_{[X=a_i]}$, $Y = \sum_j b_j \cdot 1_{[Y=b_j]}$. X and Y are called *independent* if $E(1_{[X=a_i]} \cdot 1_{[Y=b_j]}) = E(1_{[X=a_i]}) \cdot E(1_{[Y=b_j]})$ holds for all i, j.

This extension is necessary, because there exist counterexamples, such as



$E(X)=1$, $E(Y)=2/3$, $E(X \cdot Y) = 2/3$ but the information about X changes the betting on Y. Reason for this: $E(1_{[X=2]} \cdot 1_{[Y=1]}) = 1/3$ and $E(1_{[X=2]}) \cdot E(1_{[Y=1]}) = 1/3 \cdot 1/3 = 1/9$.

IX. Variance, Standard deviation

This is introduced as usual. We define $Var(X) := E((X-E(X))^2) = E(X^2) - E^2(X)$, (the second "=" holds by linearity of E), and $\sigma(X) := \sqrt{Var(X)}$.

As an example (which is referred to below) take Bernoulli trials. There are n independent copies of with E = p. Let H_n := Number of 'successes'.

$H_n = 1_{[\text{success at the first trial}]} + \ldots + 1_{[\text{success at the } n^{th} \text{ trial}]}$

$E(H_n) = \quad p \quad + \quad \ldots \quad + \quad p \quad = \quad np$

Now to the variance. $\mathrm{Var}(1_{[\text{success in the } i^{\text{th}} \text{ trial}]}) =$

$$E(1^2_{[i]}) - E^2(1_{[i]}) = p^2 - p = p(1-p) =: pq.$$

Then $\mathrm{Var}(H_n) = \sum^n \mathrm{Var}(1_{[i]}) = npq.$

X. Conditional Probability, Bayes' rule

Define (1) $P_B(A)$ (or $P(A|B)$) $:= \dfrac{P(A \text{ and } B)}{P(B)}$, $P(B) \neq 0$.

(2) $E_B(X) := \sum_i a_i \, P_B(X=a_i)$, if $W(X) = \{a_1, a_2, \ldots, a_n\}$.

(1) gives $P(A \wedge B) = P_B(A) \cdot P(B)$ and by reasons of symmetry

(2) gives $P(A \wedge B) = P_A(B) \cdot P(A)$. From these two, one has the important equality (3) $P_A(B) \cdot P(A) = P_B(A) \cdot P(B)$, which actually contains Bayes' rule. Now, if $1_{[A_1]} + \ldots + 1_{[A_n]}$ is a mutually exclusive decomp. of 1, and B is an arbitrary event, then a) $1_{[B]} = 1_{[A_1 \wedge B]} + \ldots + 1_{[A_n \wedge B]}$, trivial.

b) $P(B) = P(A_1 \wedge B) + \ldots + P(A_n \wedge B)$, apply E to a).

By definition (1): c) $P(B) = P(A_1)P(B|A_1) + \ldots + P(A_n)P(B|A_n)$

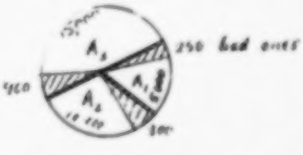
d) $P(A_i|B) = P(A_i) \, P(B|A_i) \, / P(B)$, according to (3).

which gives *Bayes' Rule* " $= \dfrac{P(A_i) \, P(B|A_i)}{P(A_1)P(B|A_1) + \ldots + P(A_n)P(B|A_n)}$

How do (should) we work with Bayes' rule? We use the spinner, viz.

Example: Three machines A_1, A_2, A_3 produce nails, 5,000, 10,000, 15,000 per day, respectively. The dross rates are 5%, 3%, and 6% respectively. In the evening one nail is taken by chance. What is the probability that it is "bad"? If it is bad, what is the probability that it comes from A_1?

The situation is modelled with the spinner. The first question corresponds to the expectation that the spinner stops in one of the hatched sectors. The second means it has stopped there and now we are interested in the chance that it is the one belonging to A_1. The solutions are found immediately:



(proportions not correct!)

$$P(\text{bad}) = \frac{300+900+250}{5{,}000+10{,}000+15{,}000} = \frac{1{,}450}{30{,}000} \approx 4.83\%.$$

$$P(A_1 \,|\text{bad}) = \frac{250}{250+300+900} = \frac{250}{1{,}450} \approx 17.24\%$$

XI. Conditional Expectation, General product rule

As $E(X \cdot Y) = E(X) \cdot E(Y)$ is not always valid, one is led to look for a similar formula, which holds in general.

(1) It is easy to show $E(X \cdot 1_{[A]}) = E(E_A(X) \cdot 1_{[A]})$. This is a special case of a general rule. To get it, define the conditional expectation. Let X, Y be RVs with standard decompositions $\Sigma\, a_i 1_{[X=a_i]}$, $\Sigma_j\, b_j 1_{[Y=b_j]}$. Then $E_Y(X) := \Sigma_j\, E(X|Y=b_j) \cdot 1_{[Y=b_j]}$ is called the *expectation of X under the condition Y.*

Up to now the condition was of type "A", fixed. Now, A may vary according to the outcomes of Y. Therefore $E_Y(X)$ is not a number, but a RV itself. With this the general product rule reads

$$E(E_Y(X) \cdot Y) = E(X \cdot Y). \qquad (2)$$

This equation holds in both cases X, Y dependent, independent (the latter is trivial). (Note: the values of $E_Y(X)$ are uniquely defined except for values b_j of Y with $P(Y=b_j) = 0$.)

This rule allows to calcualte E(XY) without discussing XY.

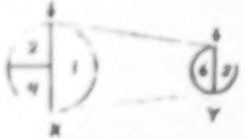
Example: X and Y are linked (e.g.) in a way that the Sp which represents Y runs two times as fast as the spinner-game X:

The dependency is expressed by the table:

	2	6
1	1/4	1/4
2	0	1/4
4	1/4	0

X

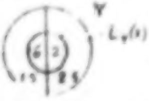
Y

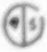


Construction of $E_Y(X)$ by calculating $E(X|Y=2)$ and $E(X|Y=6)$.

$E(X|Y=2) = 1 \cdot \frac{1/4}{1/2} + 2 \cdot 0 + 4 \cdot \frac{1/4}{1/2} = 2.5$, and similar $E(X|Y=6) = 1 \cdot 5$

These values have to be written "upon" Y:



The product $E_Y(X) \cdot Y$ looks like , $E=7$.

So we also know $E(XY) = 7$, by (2).

XII. Chebyshev's inequality, Laws of large numbers.

Usually, Chebyshev's inequality is easily derived. Nevertheless, a different proof is discussed here, which could be of interest for methodological reasons. Take X with μ and σ. Then $|X-\mu|$ and $1_{[|X-\mu| \geq \varepsilon]}$ are RVs and with real $\varepsilon > 0$ $\varepsilon \cdot 1_{[|X-\mu| \geq \varepsilon]} \leq |X-\mu|$ holds (check the truth-values of both sides.

As $1^m_{[A]} = 1_{[A]}$ we obtain $\varepsilon^m \cdot 1_{[|X-\mu| \geq \varepsilon]} \leq |X-\mu|^m$, $m \in \mathbb{N}$.
Dividing by ε^m and taking expectations (1) $P(|X-\mu| \geq \varepsilon) \leq \frac{E(|X-\mu|^m)}{\varepsilon^m}$
which contains *Chebyshev's inequality* ($m=2$).

Note: The proof holds as well if X is a continuous RV.

Now, *Bernoulli's theorem* is soon derived by (1). Remember the example in IX and define $h_n := H_n/n$. Then $\sigma^2_{h_n} = pq/n$, $E(h_n) = p$. Inequality (1) and the identifications $X=h_n$, $\mu=p$, $m=2$, give $P(|h_n - p| \geq \varepsilon) \leq \frac{pq}{n\varepsilon^2}$, which is Bernoulli's law.

With similar arguments the weak law of large numbers for a sequence (X_n) of independent RVs with existing μ_n and σ^2_n may be derived if
$\frac{1}{n} \sum_{i=1}^{n} \sigma^2_i \xrightarrow[n \to \infty]{} 0$ is assumed to hold.

As to the strong law of large numbers take again H_n and h_n. Formula (1), $m=4$, a rather lengthy (but elementary) calculation shows that $P(\lim_{n \to \infty} h_n = p) = 1$ holds too. (More general statements need different techniques and more effort).

XIII. Probability distribution functions.

Probability distributions may be introduced as usual. But one particular hint should be added here. We have stressed several times that E applies both for independent and dependent sums of RVs. And so, the calculations of the Es of the binomial and hypergeometric distributions run quite analogously, as follows. The binomial RV gives the number of successes in n Bernoulli trials, for example X is the number of successes in a sample of n elements drawn with replacement from a population of N elements of two kinds: N_1 elements labelled "success" and $N-N_1$ labelled "failure".

The hypergeometric RV is the number of successes in a similar experiment, but now drawn without replacement. The decomposition of X and Y may be chosen as follows:

$X = 1_{[\text{success at first draw}]} + .. + 1_{[\text{succ. at } n^{\text{th}}]}$, $Y = 1_{[\text{succ. at 1st}]} + .. + 1_{[\text{succ. at } n^{\text{th}}]}$

In both cases $E(1_{[\text{success at the } i^{\text{th}} \text{ draw}]})$ equals $\frac{N_1}{N}$ ($=: p$).

So $E(X) = np$ and $E(Y) = np$. For a full discussion of this, see [4].

XIV. Statistics.

To show one important difference between probability (theory) and statistics, we present (in one of the first lessons in this course) a spinner-game which is placed in a way that the students in the class cannot see its face. Now a partition is drawn and two outcomes, 0 and 100 say, are given. The students know these values, but not the size of the sectors. How can we find the expected value? Let us play! The teacher only tells the outcomes. The longer the list, the better the insight in this game. Later this action may be "justified" in the context of the laws of large numbers.

I feel that this situation (and its generalisations) reflects typical statistical aspects.

Some specific methods which describe 'decision under uncertainty' are worth being taught. Therefore we pick up some statistical tests. I will not omit to refer to two recent papers, which are dedicated to this problem and contain hints for classroom work, see [4], [7].

One example should be mentioned here. Let me enter the "runs", where we consider a finite 0-1-sequence. The number of runs (=maximal subsequences of symbols of like kind) may serve as an indicator of the randomness of the sequence. Too many runs (or too few) are incompatible with the idea of random action. Hence, it is essential to know the expected number of runs in a random sequence. We will derive this by our standard technique, i.e. adequate decomposition and application of E.

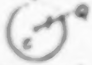
Let N, M be the number of zeros and ones, respectively, in the sequence; the total number of elements is $L = N+M$. Define a RV X which counts the number of runs and a RV Y which is the number of alternations in the sequence. Clearly $1 \leq Y \leq L-1$ and $X = Y+1$, so $E(X) = E(Y) + 1$. Determine E(Y).

$Y = 1_{[\text{alternation from the first to the second symbol}]} + \ldots + 1_{[\text{alt. from } L-1 \text{ to } L]}$.

An elementary combinatorial calculation gives
$E(1_{[\text{alt. from } i \text{ to } i+1]}) = N/L \cdot M/(L-1) + M/L \cdot N/(L-1) = 2\,\frac{NM}{L(L-1)}$, for all i.
Now $E(Y) = (L-1) \cdot 2\,\frac{NM}{L(L-1)} = 2NM/L$, and finally $E(X) = 1 + 2NM/L$.

Note: A full description of this was communicated by Ruma Falk, Jerusalem. Also she has extended the concept of a run to higher-dimensional "sequences". The calculations are similar there, because the above method works well again. See [4].

XV. Final remarks.

1) A last example of the usefulness of the spinning wheel as a model for RVs will be given with this simple game 

i.e. one will win 1 if Q comes to the pointer, and zero otherwise. What is E? The students themselves found the answer. ("If E were > 0, then we could find a sector which includes Q, but has a smaller expectation than "Q". So, E must be zero."). This simple example shows that the conclusion

P "Q") = 0 ⟶ "Q" impossible, does not hold in general. Furthermore, this problem can serve as motivation for 'infinity-problems'.

2) Most of the rules in which E is involved are valid for both discrete and continuous RV_s (e.g. Chebyshev's inequality). This is based on the properties of E : *positive linear functional*. Students may already know integrals and that these carry these properties as well. From this point of view, the use of integrals in the context of continuous RVs seems quite natural. If students do not know integrals, our concept provides an introduction of this notion. (It is easier done - with the preparation by this course - as the catchword Stieltjes-, Lebesgue-, Daniell- integral may foreshadow).

3) Combinatorics - which is an important 'pillar' in the stochastics course - is only little influenced by our concept. The notations are easily adapted, the procedure runs as usual. Maybe some exercises will lose their effects, compare Example 2, V., or the 'run-test'-derivations in $\underline{\overline{X}}$., where the distribution function of X or Y and the successive combinatorial E-calculations were bypassed.

References

[1] H.-J. Bentz and G. Palm (1980) Wahrscheinlichkeitstheorie ohne Mengenlehre, *Mathematica Didactica*, 3, 167-183.

[2] H.-J. Bentz (1980) Das Prinzip der 'getrennten Kassen' in der Wahrscheinlichkeitsrechnung, in: *Beitrage zum Mathematik-Unterricht*, Schroedel, Braunschweig.

[3] H.-J. Bentz (1982) *Zum Wahrscheinlichkeitsbegriff von Chr. Huygens*, to appear.

[4] R. Falk (1981) *Some distributions and expectations - simplified*, Preprint, Jerusalem. A German translation, entitled Vereinfachte Darstellung einiger Verteilungen und ihrer Erwartungswerte is to appear in: *Probleme im Umgang mit dem Zufall*, Der Mathematikunterricht, Klett, Stuttgart.

[5] H. Freudenthal (1973) *Mathematics as an educational task*, Reidel, Dordrecht.

[6] H. Freudenthal (1980) Huygens' foundation of probability, *Historia Mathematica*, 7, (2), 113-117.

[7] R. Ineichen (1982) Zufallig oder nicht zufallig? in: *Probleme im Umgang mit dem Zufall*, Der Mathematikunterricht, Klett, Stuttgart (to appear).

A SURVEY OF PROBABILITY CONCEPTS IN 3000 PUPILS AGED 11-16 YEARS

D R Green, University of Loughborough, England

AIMS OF THE PROJECT

The aims of this project, sponsored by the Social Science Research Council from November 1978 to October 1981, were:

a) to survey the intuitions of chance and the concepts of probability possessed by English school pupils aged 11 to 16 years.

b) to establish patterns of development of probability concepts and to relate these to other mathematics concepts, supplementing the work of the SSRC funded project 'Concepts in Secondary Mathematics and Science' based at Chelsea College, 1974-1979 (refs. 1,2).

c) to investigate pupil responses to GCE 'O' Level and CSE probability items, with particular references to sex differences.

This paper concentrates on (a). Full details of the whole project are to be found in a published monograph (Ref. 3).

TEST DEVELOPMENT

A special Probability Concepts Test was devised over a two-yearperiod (1978-80) entailing six pilot versions before achieving the definite form. The Test comprised 26 questions, subdivided into 58 items. The full test and all item facilities are published elsewhere (Ref. 3).

DATA SOURCES

a) Altogether about four thousand pupils were tested which ultimately led to a stratified sample of 2930 pupils. Children in English state comprehensive mixed schools aged 11 to 16 years (i.e. in normal secondary school Years 1 to 5) were tested on a class basis with:

i) Probability Concepts Test (about 1 hour) (Ref. 3).

ii) AH2 Test of General Reasoning Ability (about 1 hour) (Refs. 4,5).

b) Some pupils additionally were given CSMS Tests on Fractions and/or Ratio and Proportion (Ref. 1), and 44 pupils were given individual follow-up interviews after the two main tests.

SAMPLE DETAILS

From the original four thousand subjects, those present for one test only were eliminated and from the remainder a stratified sample was drawn, to reflect the proportions of the various intellectual levels in the school population as a whole. This was based on the AH2 grade distribution which is purported to be as follows (Refs. 4,5).

A	B	C	D	E
Top 10%	Next 20%	Middle 40%	Next 20%	Bottom 10%

The final 'Sample' (N = 2930) was as indicated in Table 1.

TABLE 1: Sample numbers by School Year and by AH2 Grade

School Year	A	B	C	D	E	Overall	Mean age on taking AH2 test
1	64	128	256	128	64	640	11.7
2	67	134	268	134	67	670	12.7
3	67	134	268	134	67	670	13.7
4	54	108	216	108	54	540	14.7
5	41	82	164	82	41	410	15.7
TOTAL	293	586	1172	586	293	2930	

All the participating schools were selected from volunteers, obtained by writing to all relevant schools within 50 miles of Loughborough, England.

TESTING PROCEDURE

In most cases pupils took the Probability Concepts Test first and one or two weeks later took the AH2 Test of General Reasoning Ability. (The mean time lags between the tests for years 1 to 5 were respectively 11, 17, 12, 16, 15 days). Most testing was carried out from September to December 1980, although some took place in January to March 1981.

In order to minimise possible reading difficulties, and to ensure that pupils attended properly to each of the 26 questions of the Concepts Test, and to prove comparable conditions to all classes tested, it was decided that:

i) The test would be administered by members of the project team, and preceded by a prepared introduction explaining its nature and purpose.

ii) The test would be read out to all 1st, 2nd and 3rd year classes (ages 11-14), total time being 48 minutes.

iii) The reading out would be so controlled as to give the same amount of time to a question for all pupils.

iv) The Test would be read out to any 4th and 5th year classes (ages 14-16) for which this was deemed advisable by the class teacher. Alternatively, adequate time would be provided for silent working (55 minutes).

PROBABILITY CONCEPTS TEST - INDIVIDUAL QUESTIONS

In this paper just five of the twenty-six questions comprising the Concepts Test will be briefly discussed.

Question 1

A small round counter is red on one side and green on the other side. It is held with the red face up and tossed high in the air. It spins and then lands. Which side is more likely to be face up, or is there no difference? Tick the correct answer:

(A) The red side is more likely ☐
(B) The green side is more likely ☐
(C) There is no difference ☐
(D) Don't know ☐

Year	A	B	C*	D	Omits	Errors
1	18	28	45	9	0	0
2	15	24	55	6	0	0
3	11	20	66	2	0	0
4	12	13	71	4	0	0
5	11	12	74	4	0	0
All	14	20	61	5	0	0

Note: An asterisk (*) indicates the most appropriate response.

The facility levels for this apparently trivial question indicate that many children aged 11-16 have a very wayward probability intuition. There is some evidence of the phenomenon of 'Negative Recency' operating. It was hypothesised that by emphasising that the counter was red face up a larger number would opt for green being more likely next time than would opt for red. This has clearly been shown for Years 1, 2, 3. The phenomenon virtually disappears in the transition from Year 3 to Year 4, a surprising result by itself.

Question 6

This question comprised five items concerning 'comparison of odds' in the familiar context of coloured counters in bags. The items and the responses are summarised in Table 2. In each case Black was the required colour.

	First Bag		Second Bag			
Item	Black	White	Black	White	Diagram given	Overall facility
6a	3	1	2	1	Yes	88
6b	5	2	5	3	Yes	67
6c	2	2	4	4	Yes	60
6d	12	4	20	10	No	57
6e	3	1	6	2	No	33

TABLE 2: Details of items concerning comparison of odds

To show the nature of the items, 6(d) is presented in full:

Question 6(d)

Two other bags have black and white counters:

Bag G: 12 Black and 4 White
Bag H: 20 Black and 20 White

Which bag gives a better chance of picking a black counter?

(A) Same chance
(B) Bag G
(C) Bag H
(D) Don't know

Why? ....

Year	A	B*	C	D	Omits	Errors
1	8	38	51	2	0	0
2	6	50	41	4	0	0
3	6	62	31	1	0	0
4	5	68	26	0	0	0
5	4	74	20	2	0	0
All	6	57	35	2	0	0

Analysis of both verbal explanations and multiple choice responses was undertaken to throw light on the strategies employed by pupils in solving these items. Although many theoretical strategies can readily be imagined the incidence of any one strategy being consistently applied to the five items in Question 6 is small. The commoner identifiable strategies are indicated in Table 3.

Strategy Number	Strategy Description	Frequency of consistent application	Percentage of Sample
1	Choose the bag with more total counters	41	1.4
2	Choose the bag with more Black counters	221	7.5
3	Choose the bag with greater Black-White difference	92	3.1
4	Choose the bag with greater Black:White ratio	583	19.9
		937	32.0

TABLE 3: Frequency of consistent responses for various theoretical strategies

It is clear that relatively few pupils aged 11-16 years (32%) consistently apply any one pure strategy to all five problems. By far the commonest single pattern observed is for the correct proportional reasoning strategy (20%). This analysis immediately indicates how unlikely it is that a straightforward 'model' of pupil behaviour could be devised. Various strategies are invoked depending on the circumstances. This point was brought out very clearly in individual interviews - for Question 6(a) the almost unanimous reason given was simply "Because A has more Blacks". Only with great perseverance could the investigator occasionally evoke a reason employing the ratio concept, even from pupils who readily supplied just such a reason for 6(d) or 6(e) where such a concept is essential.

A more detailed Year by Year breakdown is provided in Table 4.

Strategy Number	Strategy Description	Perfect pattern frequency as % of Year				
		1	2	3	4	5
1	More counters (Total)	2	2	1	0	0
2	Comparing BLACK (Counting)	13	10	6	4	3
3	B-W (Difference)	6	2	3	3	1
4	B:W (Ratio)	7	11	21	28	42

TABLE 4: Incidence of strategy response patterns by Year

The growth with age of the usage of the Ratio concept (4) and the decline of counting (2) and Differencing (3) is pronounced. Differencing would appear to run parallel to Counting rather than supersede it as the pupils grow older. The growth of the Ratio strategy is on a scale much greater than can be attributed to its replacing Counting and Differencing.

Table 5 shows several other popular response patterns.

Response Pattern	Response Pattern						
Number	6(a)	6(b)	6(c)	6(d)	6(e)	Frequency	%
(i)	A	A	C	B	B	82	2.8
(ii)	A	A	C	B	C	316	10.8
(iii)	A	A	C	C	A	72	2.5
(iv)	A	A	C	C	C	214	7.3
(v)	A	A	B	B	C	93	3.2

(Underlined responses are correct)

TABLE 5: Popular response patterns not corresponding to a theoretical strategy

Response patterns (i) and (ii) are correct apart from item 6(e) and their relative frequencies show the clear preference for choosing the bag with more black counters (C instead of B). Patterns (iii) and (iv) both are in error by wrongly selecting for item 6(d) the bag with more black counters, and that is the flaw manifested in pattern (v) also. Together, these five common patterns represent about 27% of the Sample. The inference is that many pupils use a mixture of strategies, competently handling problems where only one colour varies in number, but being less sure in cases where both colours' numbers vary, and generally reverting to selecting the bag with more black or greater black/white difference. This would account for perhaps 50% of the population aged 11-16 years. More detailed analysis surprisingly showed that, with the exception of the development of the correct Ratio strategy, percentages were very nearly static with increasing age.

Also, an analysis by AH2 grade was undertaken. It can be seen from Tables 6 and 7 that use of the Ratio concept is very much a function of intelligence, more so than of age - compare with Table 4.

	AH2 Grade				
Strategy	A	B	C	D	E
1 (Total, more)	0	1	1	2	3
2 (Comparing Black)	1	5	8	12	10
3 (B-W difference)	0	2	3	5	5
4 (B:W ratio)	54	33	16	5	2
1 to 4 combined	55	41	28	24	20

TABLE 6: Percentage of pupils in each Intelligence level whose response patterns match a theoretical strategy

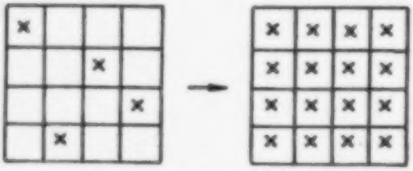
	AH2 Grade				
Year	A	B	C	D	E
1	31	13	4	1	0
2	36	23	6	1	0
3	64	36	15	5	0
4	70	45	25	7	6
5	83	62	43	17	7

TABLE 7: Percentage of pupils in each Intelligence level for each Year whose response patterns correspond to the Ratio strategy

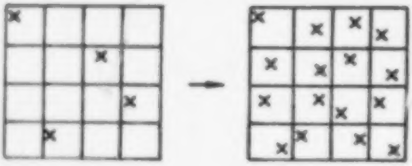
Question 11

The flat roof of a garden shed has 16 square sections. It begins to snow. At first just a few snowflakes fall, then after a while more have landed. Below are three sets of two pictures. Each set shows the pattern of snowflakes building up - first 4 flakes, then 16 flakes.

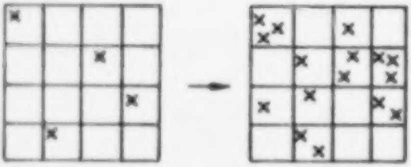
SET A



SET B



SET C



QUESTION: Which of these sets best shows the kind of pattern you would expect to see as the snowflakes land?

Set A ☐
Set B ☐
Set C ☐
Sets B and C ☐
Each kind of pattern is as likely ☐

Year	A	B	C*	B and C	'equal'	Omits	Errors
1	11	12	26	23	27	0	1
2	13	12	26	20	29	0	1
3	13	14	24	25	24	0	0
4	13	16	16	21	33	0	0
5	12	15	18	21	34	0	0
All	12	14	23	22	29	0	0

The astonishing thing about this item is that performance declines with age. These rsults confirm what had been found in Pilot Testing as reported earlier by Green (Ref. 6). It can be hypothesised that we see at work here two opposing tendencies - maturation/experience on the one hand and dominance of mathematical/scientific deductivism on the other which stifles the appreciation of randomness by seeking to codify and explain everything.

Question 20

A teacher asked Clare and Susan each to toss a coin a large number of times and to record every time whether the coin landed Heads or Tails. For each 'Heads' a 1 is recorded and for each 'Tails' an 0 is recorded.

Here are the two sets of results:

CLARE:
01011001100101011011010001110001101101010110010001
01010011100110101100101100101100100101110110011011
01010010110010101100010011010110011101110101100011

SUSAN:
10011101111010011100100111001000111011111101010101
11100000001000101001000001000110001010000000011001
00000001111100001101010010010011111101001100011000

Now one girl did it properly, by tossing the coin. The other girl cheated and just made it up.

(a) Which girl cheated? ANSWER ....

(b) How can you tell?

ANSWER ....

....

This question was modelled on one suggested by M. Halmos (Ref. 7).

20(a)

Year	Clare*	Susan	Cannot tell	Omits
1	38	53	1	7
2	34	56	2	7
3	30	59	4	7
4	40	46	3	9
5	35	49	5	9
All	35	53	3	8

There is no improvement with age, and for all Years more pupils incorrectly conclude that Susan cheated rather than Clare.

Analysis of the reasons given is also interesting:

20(b)

	Facility	(1) Pattern Regularity	(2)	(3) Inequality of O, I.	(4)	(5) Equality of O, I.	(6)	(7) Run Lengths	(8)	Other	Omits
Year	(1)+(7)	C	S	C	S	C	S	C	S		
1	21	20	2	8	21	2	1	1	15	17	13
2	21	17	3	7	21	1	0	4	18	13	14
3	23	21	4	4	17	3	0	2	21	14	14
4	28	26	3	3	14	4	0	2	20	16	11
5	28	24	6	1	16	4	0	4	19	12	12
All	24	21	3	5	18	3	1	3	18	14	13

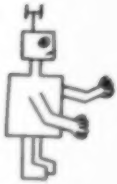
The reasons were generally of the following type, corresponding to the above abbreviated headings:

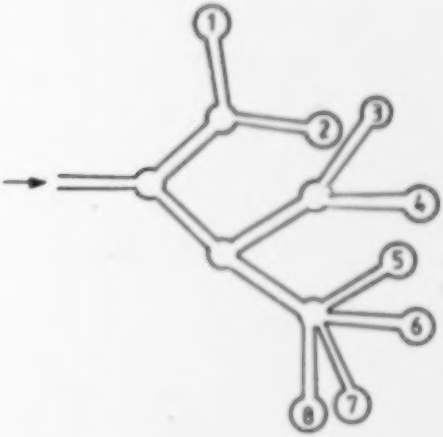
1) Clare too regular pattern (21%)
2) Susan too irregular pattern (3%)
3) Clare too many Is (or Os)/too far from 50-50 proportion (5%)
4) Susan too many Os (or Is)/too far from 50-50 proportion (18%)
5) Clare too close to 50-50 proportion (3%)
6) Susan too close to 50-50 proportion (1%)
7) Clare too short runs (3%)
8) Susan too long runs (18%)
Others + omits (28%)

Only (1) and (7) are valid reasons. Thus the true overall facility of this item is about 24%, and shows only slight improvement with age. Very few pupils thought Susan's pattern too irregular (3%) but their concern is shown in stating that her run lengths are too long (18%).

Question 22

A robot is put into a maze, which it begins to explore. At each junction the robot is as likely to go down any one path as any other (except it does not go back the way it came). There are eight traps at the ends of the eight paths (see picture). In which trap or traps is the robot most likely to finish up, or are all traps equally alike?





ANSWER: ....

Year	'1 and 2'*	1/2	5/6/7/8	'Same'	'Other'	Omits
1	5	7	22	57	6	3
2	4	5	20	62	6	2
3	7	3	19	62	7	3
4	9	7	16	59	7	3
5	13	4	15	59	5	3
All	7	5	19	60	6	3

This item, it was anticipated, would prove difficult for Years 1 to 3 (or 4) but would be considerably less difficult for Year 5 pupils who had met tree diagrams and multiplication of probabilities. However, in the event the item was extremely difficult for all years, the percentage considering all traps equally likely remaining virtually fixed at around 60% at each age level.

Even when treating correct a choice of either Trap 1 or Trap 2 singly the facility does not exceed 17%. Interviews with pupils showed that knowledge of tree diagrams did not seem to help. Pupils who opted for "all are equally likely" often said that "all have a chance" and "no one can say which" were phrases which implied equality.

The abysmal failure of all Years on this item is in marked contrast to Fischbein et al's results (Ref. 8, appendix 11) in an equivalent but practical setting where facilities in the range 58% to 69% were reported.

PROBABILITY CONCEPTS TEST OVERALL RESULTS

Pupils' responses to the Concepts Test have been analysed by Year, Sex, Intelligence (AH2 grade) and Mathematical Ability ("MA"). As a further step each of the 26 questions was allocated to one of three sub-test categories: Verbal (VB), Combinatoric (CB), Probabilistic (PB).

Each item attracted a mark of 1 if correct and 0 if incorrect, yielding three sub-scores (maxima VB: 15, CB: 6, PB: 29) and a total Test Score (maximum 50). The Test Score breakdowns are shown in Table 8.

	TEST SCORE		VB SCORE		CB SCORE		PB SCORE	
Year	Boy	Girl	Boy	Girl	Boy	Girl	Boy	Girl
1	20.3	19.1	7.2	7.0	2.2	2.2	10.8	9.9
2	22.4	21.2	8.1	7.9	2.4	2.5	11.9	10.7
3	25.5	24.5	8.8	8.9	3.1	3.1	13.6	12.6
4	27.3	25.5	9.7	9.2	3.2	3.3	14.4	13.1
5	28.6	27.4	10.0	9.9	3.6	3.5	16.1	14.1

TABLE 8: Mean Test Scores broken down by School Year and Sex

Analysis of Variance shows that age, intelligence and sex are significant factors, the variation attributable to each being:

AH2 Grade: 44%, School Year: 14%, Sex: ½%.

The dominance of intelligence (AH2 Grade) is clear, its influence being three times that of age (School Year). With MA substituted for AH2 Grade a similar result was obtained (40% variance), indicating that MA is almost as good a guide to Test performance as is AH2 Grade. This is not surprising in view of the high correlation (+.72) between the two factors.

PROBABILITY CONCEPTS

In order to attempt to allocate each pupil unequivocally to a concept 'level', Guttman Scalogram Analysis was undertaken. The aim was to group together selected Concepts Test questions into subtests representing different levels, and to allocate a passmark for each level. To be at a particular level a pupil must pass each lower level and also pass the given level. The main outcome was a scale comprising three levels (plus zero level indicating total 'failure'). Unfortunately, but not surprisingly, many of the more unusual and interesting items in the Concepts Test could not be fitted into these scales! The distribution of Concept Levels is shown in Table 9.

		Probability Concept Level				
Year	Sex	0	1	2	3	Mean Level
1	Boy	15	51	31	3	1.21
	Girl	18	60	21	2	1.07
2	Boy	12	44	38	6	1.37
	Girl	11	55	27	6	1.28
3	Boy	5	31	48	16	1.75
	Girl	6	43	40	12	1.57
4	Boy	6	25	44	25	1.87
	Girl	6	34	40	20	1.74
5	Boy	2	16	50	32	2.11
	Girl	3	27	44	26	1.93

TABLE 9: Percentages of Main Sample Pupils in each Concept Level by School Year and Sex

A steady progression of Concept Level with increasing age is apparent. Boys show superiority to girls consistently, but the gap between them, as judged by the mean Concept Level, does not widen between the ages of 11-16 years. Further analysis by intelligence level throws light on the nature of the differences found (Table 10). The general trend is that whereas the more able pupils gain about 1 level during the five-year period, the least able, starting 1 level behind, fall even further behind during the same period, only gaining about ½ a level.

	AH2 Grade				
Year	A	B	C	D	E
1	1.84	1.52	1.08	0.80	0.62
2	2.15	1.77	1.30	0.86	0.67
3	2.54	2.12	1.65	1.29	0.67
4	2.67	2.27	1.89	1.25	0.77
5	2.83	2.44	2.10	1.50	1.10
All	2.36	1.98	1.55	1.10	0.74

TABLE 10: Mean Concept Level by Year and by General Reasoning Ability

CROSS-REFERENCE TO PIAGETIAN STAGES

By making a detailed study of pupils' responses to the 'comparison of odds' items (Question 6), and relating this to Piaget and Inhelder's own work (Ref. 9) it was possible to indicate the ages at which Piagetian Stages are normally achieved (Table 11). It must be concluded that most English school children do not achieve the level of formal operations, which is Piaget's Stage III or the Probability Concept Level 3, a result in line with findings on Science concepts reported by Shayer, Kuchemann and Wylam (Ref. 10).

Piagetian Stage	Corresponding Question 6 Item	AH2 ability band		
		Top 30%	Middle 40%	Bottom 30%
IB	(a)	<11	<11	11
IIA	(b)	11	13	16
IIB	(c)	12	14	-
III	(d) (e)	14	-	-

TABLE 11: Ages at which Piagetian Stages are normally achieved for comparison of odds problems, for three intelligence levels

SUMMARY OF FINDINGS RELATING TO THE PROBABILITY CONCEPTS TEST (N=2930)

1) Nearly all test items show an increase in facility with age and intellectual ability.

2) Often years 1 and 2 display quite similar response patterns, with substantial increases in performance for Years 3, 4 and 5.

3) Items which merely require the counting concept are well done by all Years, but those requiring the ratio concept are very poorly done, particularly below Year 4.

4) Various strategies are employed by pupils in answering comparison of odds problems. Counting and differencing are two popular strategies, but are rarely applied consistently. Computing ratios is markedly a function of intelligence and also of age.

5) Items involving an appreciation of randomness, the stability of frequencies and inference were particularly poorly done, with little improvement with age evident.

6) Although most pupils coped well with a combinatoric item at a concrete level (permutation of 3 objects: 75%) few could proceed to formal level items (permutations of 4 objects: 12%, 5 objects: 4%).

7) Pupils' verbal ability is often inadequate for accurately describing probabilistic situations.

8) Very often certainty and high probability are equated, as are impossibility and low probability. Pupils freely associated a 50% probability with (i) outcomes which might or might not happen, (ii) equi-probable outcomes (when more than 2 outcomes exist).

9) Generally, boys outscore girls on total Concepts Test Score, but not on combinatoric questions.

10) Generally, boys show superior performance to girls on many Concepts Test questions, with girls significantly superior on one question only (Question 20).

11) General reasoning ability (AH2 Grade) is the dominant factor in Test Score (44%), being three times as important as age (School Year) (14%). Sex is of minor significance (½%).

12) Concepts Level means rise consistently with increasing age (School Year) and intellectual level (AH2 Grade).

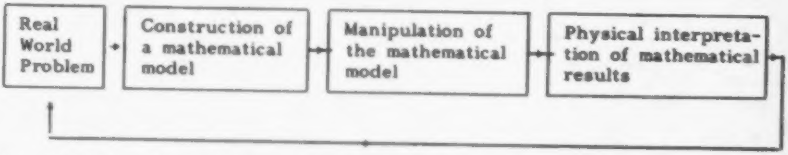
13) Generally, boys have a higher Concepts Level than girls.

14) Most pupils do not attain the level of 'formal operations' (Piagetian Stage III) by Year 5, and presumably leave school at 'concrete operations' level (Stage II).

15) Teaching style, along the 'teacher centred ... individualised learning' continuum does not appear to be a significant factor affecting Test Score.

SOME RECOMMENDATIONS

a) The subject of probability is a very appropriate topic whereby to emphasise the nature of mathematical modelling:


Real World Problem
Construction of a mathematical model
Manipulation of the mathematical model
Physical interpretation of mathematical results

Question 1 is asking the pupil what he thinks is an appropriate model, which dictates his choice of response. (Is colour a relevant factor? Is past history?).

School mathematics very much emphasises mathematical manipulation and gives very little weight to either model building or interpretation. Probability teaching must attempt to redress the balance. (Does Question 1 only have one defensible answer?).

b) Individual interviews have confirmed that some pupils are quick to answer questions intuitively, with no obvious intervention of mathematical processes. However, it is unwise for the teacher to assume that all pupils can see what to the teacher appears obvious, even if some in the class effortlessly provide the 'correct' answers. Only an extensive programme of class-based activities is likely to provide the weight of evidence and experience to eliminate the fallacious thinking exhibited by pupils and adults alike.

c) The provision of diagrams seems to help pupils' understanding, and should be provided wherever possible, or pupils encouraged to interpret verbal problems pictorially before proceeding to solve them.

d) Various strategies are employed by pupils in solving comparison of odds questions but little consistency can be found. Much benefit can accrue from explicit classroom discussion of appropriate strategies for particular problems with a view to developing adequate discernment of types of problem or of mastery of a universally applicable strategy (the ratio concept).

e) The teacher needs to give attention to educating pupils in the normally accepted usage of various terms which indicate varying degrees of likelihood. Commonality of meaning is obviously most desirable, and should not tacitly be assumed. Discussion of events incorporating a chance element needs to be encouraged.

f) In our scientifically and technologically dominated society, with its deterministically oriented education system, it is perhaps inevitable that the concept of randomness should be neglected. Very little time or attention is given to this concept despite the fact that we are surrounded by random occurrences every day of our lives. It is unreasonable to treat so lightly such a pervasive property of our world.

g) The multiplicative principle is poorly understood, and pupils have little in the way of basic intuition on which it can be based. Response to Question 22 was very poor. Clearly little conceptual understanding has been achieved, and using a tree diagram was just a mechanical device.

h) Drawing inference from the outcomes of sampling experiments and appreciation of the stability of frequencies are two most important conceptual areas. Very few pupils were able to deal with either. Our results, and follow-up interviews, suggest that much practical activity and discussion is needed to develop an adequate appreciation of these concepts in the majority of pupils.

References

[1] Hart, K.M. (1980) Secondary School Children's Understanding of Mathematics (CSMS). Chelsea College, University of London.

[2] Hart, K.M. (1981) Children's Understanding of Mathematics: 11-16, John Murray.

[3] Green, D.R. (1982) Probability Concepts in 11-16 Year Old Pupils. Centre for Advancement of Mathematical Education in Technology, Loughborough University of Technology. 40 pages, £1.00.

[4] Heim, A.W., Watts, K.P. and Simmonds, V. (1974) AH2 and AH3: Parallel Tests of Reasoning, Brit.J.Psychol., 65, 1-11.

[5] Heim, A.W., Watts, K.P. and Simmonds, V. (1978) AH2/AH3 Manual, (2nd Edition) NFER.

[6] Green, D.R. (1979) The Chance and Probability Concepts Project, Teaching Statistics, 1, 3, 66-71.

[7] Halmos, M. (1975) First steps in teaching probability: in Statistics at the School Level, edited by L. Rade, Almqvist & Wiksell.

[8] Fischbein, E. (1975) The Intuitive Sources of Probabilistic Thinking in Children, D. Reidel.

[9] Piaget, J. and Inhelder, B. (1975) The Origin of the Idea of Chance in Children, Routledge and Kegan Paul.

[10] Shayer, M., Kuchemann, D.E. and Wylam, H. (1976) The Distribution of Piagetian Stages of Thinking in British Middle and Secondary School Children, Brit.J.Educ.Psychol., 46, 164-173.

MISCONCEPTIONS OF PROBABILITY, SYSTEMATIC AND OTHERWISE; TEACHING PROBABILITY AND STATISTICS SO AS TO OVERCOME SOME MISCONCEPTIONS

J M Shaughnessy, Oregon State University

Suppose you are teaching an introductory course on probability and statistics to first year college students or to fourth year secondary students. After several weeks, when you have covered some introductory material, you present these problems to your students. What answers will your students give for the questions?

1. The probability of having a baby boy is about 1/2. Which of these sequences is more likely to occur for having six children?

   i) a) BGGBGB b) BBBBGB c) about the same chance for each.

   ii) a) BGGBGB b) BBBGGG c) about the same chance for each.

2. Consider the grids below.

Grid A	Grid B
X X X X X X X X	X X
X X X X X X X X	X X
X X X X X X X X	X X
	X X
	X X
	X X
	X X
	X X
	X X

Are there

a) more possible paths* in grid A?

b) more possible paths in grid B?

c) about the same number of paths in each grid?

(* A path is a polygonal chain of line segments, starting at top row and meeting one and only one symbol in each row. Examples have been shown to your students).

3. The chance that a baby is born a boy is about 1/2. Over an entire year, would there be more days when at least 60% of the babies born were boys in

a) a large hospital
b) a small hospital
c) makes no difference

4. You must select committees from a group of 10 people. Would there be:

a) more distinct possible committees of 8
b) more distinct possible committees of 2
c) about the same number of committees of 8 as committees of 2?

5. There are two cab companies, Green Cab and Blue Cab, in a city. Of all the cabs, 85% are Green and 15% are Blue.

A cab is involved in an accident, and a witness testifies that the cab was Blue. Past data on the accuracy of witness identification indicates that witnesses make correct colour identifications in 80% of the cases, and incorrect in 20% of the cases.

What is the probability that the cab in the accident was Blue rather than Green?

Of course, none of these questions is the type students in the first few weeks of an introductory course would be expected to do by applying theoretical principles. The point is: how do statistically naive students respond to such questions? Naive students would likely possess certain misconceptions of probability due to their inexperience with the mathematical laws of probability. Thus, it may be possible to clear up some misconceptions by familiarising students with probability concepts and theory. However, in the research that has been reported on these and similar problems, (Kahneman and Tversky 1972, 1973, 1974; Shaughnessy 1977, 1981; Bar-Hillel 1980; Tversky and Kahneman 1971, 1980; Well, Pollatsek and Konold 1982) the evidence suggests that some misconceptions of probability are rather deep seated and may not be the result of simple mathematical naivete. Some misconceptions appear to be of a psychological kind, and more exposure to the laws of probability and statistics may not be sufficient to overcome them, even in the best of courses. Indeed, some of these "non-mathematical" misconceptions appear to be systematic, that is, they can perhaps be characterised by certain psychological principles. On the other hand, some misconceptions may be very unsystematic and possibly the by-product of certain "belief systems" possessed by the students.

In this paper, we will discuss the types of misconceptions that arise from students' responses to the five questions above. Then we will consider the implications of these misconceptions for the teaching of probability and statistics, and suggest some approaches to probability that may be useful for confronting the misconceptions that our students possess.

TYPES OF MISCONCEPTIONS

Errors Based on Intuitive Heuristics

Consider once again Problem 1.

(1) Which sequence is more likely to occur, BGGBGB or BBBBGB? Here are the responses from a random sample of 80 college undergraduates prior to a course in probability. The numbers are frequency of responses.

BGGBGB	BBBBGB	Same
50	2	18

A large majority of students say that the sequence BGGBGB is more likely to occur (Kahneman and Tversky, 1972; Shaughnessy, 1977). When asked to give a reason for their choice they often say "because BGGBGB fits the expected 50:50 ratio of boys to girls". The sequences are actually equally likely to occur, each with probability, 1/64. These students have used the probability of a boy one try, 1/2, as an indication of the likelihood of a whole sequence of trials. Kahneman and Tversky refer to this strategy, or intuitive heuristic, as representativeness. According to respresentativeness, subjects tend to estimate likelihoods based on how similar an event is to the distribution from which it is drawn or upon how similar the event is to the process by which the outcomes are generated. Thus, BGGBGB is chosen by subjects to be more likely than either BBBBGB or BBBGGG. The former is not as "representative" of a 50-50 distribution as BGGBGB, and the latter does not reflect (represent) the random process of having children as well as BGGBGB.

The use of representativeness by subjects is also apparent in Problem 3, in a somewhat different manner.

(3) Are there more days during the year when at least 60% of the babies born are boys in a large hospital or in a small hospital?

Small	Large	No Difference
17	15	48

Many students felt that the distribution of boys in the daily sample had nothing to do with hospital size. However, there is a good chance that at least 60% of the babies are boys on a given day if the hospital has only 3 babies that day! The students used the 50:50 distribution of boys to girls in the population as a representative estimate, regardless of the sample size. Kahneman and Tversky (1972) suggest that the use of representativeness as a heuristic for estimating likelihoods can lead to errors that arise from

1. Insensitivity to prior probabilities and disregard for population proportions.
2. Insensitivity to the effects of sample size on predictive accuracy.

3. Unwarranted confidence in a prediction that is based upon invalid input data.
4. Misconceptions of chance, such as the gambler's fallacy (coin with memory).
5. Misconceptions about the tendency for data to regress to the mean.

Another systematic heuristic that is used to estimate likelihoods is availability. When people make predictions based on how accessible instances of an event are to memory, or on how easy it is to construct particular instances of the event, they are relying upon availability. For example, suppose a person is asked to estimate the percentage of smokers in her community. If she has frequent contact with people who smoke, her estimate will undoubtedly be much higher than an estimate by a person who seldom socialises with smokers. The ease with which positive instances of an event can be called to mind greatly influences estimates for the likelihood of an event.

Consider our introductory problems 2 and 4.

(2) Which grid has more paths?

A. X X X X X X X X X
X X X X X X X X X
X X X X X X X X X

B. X X
X X
X X
X X
X X
X X
X X
X X
X X

Here are the results from the sample of 80 college students.

A	B	Same
53	8	15

Although it appears easier to construct instances of paths in grid A - there seems to be more choices available at each line - there are actually 512 distinct paths in each grid since $2^9 = 8^3 = 512$. When asked to give a reason why they chose grid A over grid B, students often said "It's easier to draw a path in grid A". Results of this question strongly suggest that the heuristic of availability has influenced many of the responses.

(4) Are there more distinct committees of 2 or of 8 that can be chosen from a group of 10 people?

Here are the responses.

Committees of 8	Committees of 2	Same
7	47	19

It is easier for people to generate a lot of different committees of 2 quickly than to create committees of 8. Thus, there would seem to be more committees of 2 because they are more "available". Most combinatorially naive people do not see the dual nature of the problem, that a committee of 8 is the same as a non-committee of 2.

It is perhaps not too surprising that combinatorially naive people, such as those in the sample above with little training in probability or statistics, would rely upon some intuitive heuristics for events. However, Kahneman and Tversky (1971) have reported that subjects with a substantial background in probability and statistics were subject to the same intuitive misconceptions about probability as the naive college students. This suggests that mere exposure to probability and statistics in course work may not be enough to overcome such systematic biases as those introduced by heuristics such as availability and representativeness. In fact, the use of such heuristics may often be perfectly justified and even beneficial. If you have seen several automobile collisions at a certain dangerous intersection, the use of availability to estimate the safety of the corner may make you cautious and save an accident. A doctor may be able to make a quick diagnosis because "there is a lot of that going around", again using availability as a guessing heuristic. Of course, there's always the unfortunate instance when this backfires, when the symptoms are the same but the diseases aren't! The challenge to those of us who teach probability and statistics is to sensitise our students to the existence of such intuitive heuristics in such a way that they may determine when they are useful and when they may be misleading.

Errors Arising From Neglect of Base Rate Data

There has been a considerable amount of research on the taxicab problem (number 5) given in the introductory section above (Tversky and Kahneman, 1980; Well et al., 1982). This problem, and similar ones, have been used to determine how subjects who are unfamiliar with probability make estimates for likelihoods when several pieces of information have to be integrated.

Consider the problem again.

(5) 85% of the cabs are Green and 15% are Blue. A cab has an accident. A witness who is accurate 80% of the time identifies the cab as Blue. What is the probability it was Blue rather than Green?

Tversky and Kahneman (1980) report that the median response among several hundred subjects was .8. They suggest that subjects tend to ignore the given base data of 85% Green cabs when making their decision. Furthermore, they argue that people will be more likely to integrate base rate data into their estimate if a "causal" influence is introduced into the base rate, for example by changing the problem to read, "...85% of accidents involving cabs involve Green cabs and 15% involve Blue cabs". The actual probability that the cab was Blue (B) given that it was identified as Blue (b) is

$$P(B/b) = \frac{P(B\wedge b)}{P(B\wedge b)+P(G\wedge b)} = \frac{.12}{.12+.17} = .41$$

Interest in these types of problems arose when Edwards et. al. (1963) suggested that people may make estimates for the likelihood of events according to Bayes' Theorem, only in a more conservative way than the theorem would predict. However, the results of Kahneman and Tversky imply that people do not act as intuitive Bayesians, at least not on the taxicab problem.

Recent research by Well et. al. (1982) suggests that the types of misconceptions that occur among naive subjects when confronted with Bayesian problems, such as the taxicab problem, are much more complicated and varied than simple neglect of base rate data or use of some causal schemas. Much of the research on misconceptions of probability has reported only average responses - means or medians - and has not asked the studetns why they responded in a particular way. Well carried out follow-up clinical interviews with students who had been given Bayesian problems, like the taxicab problem, in order to disclose their mental processes. He found large individual differences in the types of strategies employed and in the kind of misconceptions that were apparent among the students. In particular, he found no evidence to support the causal schemas contention of Tversky and Kahneman. Rather, some students used non-probabilistic strategies to estimate the likelihood of events, strategies which may be based on a deterministic model or even on personal belief systems. Well says, "It is difficult to infer a strategy from a numerical answer. We have observed subjects with markedly different strategies (in the interviews) arrive at the same numerical answer, and subjects with the same strategy arrive at different answers".

Errors Due to a Deterministic Model or Belief System

An estimate for the likelihood on an event may not be based upon a probabilistic model at all. For example, suppose someone decides that the witness is right in the cab problem. They will then conclude that the Blue cab was involved in the accident. There is no chance involved here, and no use of the percentages before the decision. In the interviews Well and his colleagues obtained estimates as high as 90-95% that it was a Blue cab. When asked why, these subjects said, "I think the witness was right", or "I think it was a Blue cab". In such cases, the decision is based on a belief or deterministic decision. The actual percentage given there is more like a confidence rating than a probability estimate. Base-rate information may be neglected if it violates preconceptions or beliefs of the subject.

This type of misconception may pose the most difficult challenge of all to those of us who teach probability and statistics. What do you do if your students just plain don't believe in probability or chance? Well (1982) reports a protocol excerpt in which a student is asked to pick which side he thinks will turn up if an irregularly shaped die (actually a bone) is rolled. During the interchange the student eventually settles on side C as the most likely to come up - for no apparent probabilistic reason. (In fact, there is no evidence of an understanding of chance by the student during the entire episode). The student is shown data from 1,000 rolls of the die which clearly suggest side D is most likely to occur, and then asked which would be the most likely. "You found D most likely. ... I still think C". This student has evidently formed a belief that C is most likely, and experimental data are not going to alter the belief.

A PROBLEM FOR TEACHERS OF PROBABILITY

Most of the literature referenced above comes from research on judgement and decision making by psychologists. However, this research has very serious implications for the teaching of probability and statistics. The student in our introductory courses, either in secondary school or in college, come already armed with a set of preconceptions and misconceptions about probability. The misconceptions may be of a systematic probabilistic nature such as representativeness and availability, or they may be non-probabilistic in nature, based on causal schemas, belief systems, or other deterministic attitudes. The intuitive probabilistic heuristics may be so ingrained that they are difficult to overcome In a course on probability and statistics, even one which is especially designed to confront them (Shaughnessy, 1977). The deterministic misconceptions of the students may pose a more serious problem for the teaching of probability. The teacher may have to break through a belief barrier to even reach the students in the first place. Fischbein (1976) claims that science education traditionally has emphasized deterministic models and has tended to neglect probabilistic models and the study of uncertainty. Thus our intuition for probabilistic thinking may be distorted by an overemphasis on deterministic models in our schools, such as natural selection, Newton's laws of motion, or a deductive axiomatic approach to geometry. Can we then be too surprised at the deterministic reasoning of some of our students when we finally do get them in a course on probability?

What kinds of things should we do in our probability and statistics courses in order to confront and perhaps overcome some of these misconceptions? I do not claim to have a definitive answer to this question, but I do believe there are some things that will help. These suggestions certainly are not brand new, but I feel they are all the more important when one considers the misconceptions of probability indicated by psychological research.

TEACHING PROBABILITY TO OVERCOME MISCONCEPTIONS

Let us consider some recommendations which may help avert or overcome our students' misconceptions or probability and statistics.

1. Introductory probability and statistics should be activity-based and experimental. We can proceed from guesses to experiments to simulations of experiments, and only after considerable intuition has been developed, to a formal mathematical presentation. Confront students with problems in which misconceptions can arise, such as the introductory problems in this paper.
2. Emphasize the confounding role that belief systems can play in judgement and decision making. At the same time, teach the importance and usefulness of statistics for helping to make objective judgements and decisions.
3. Sensitize students to misuses of statistics.
4. Give students challenge problems which may require them to gather or simulate their own data, and then make a decision (i.e., synthesize 1-3 above).
5. Sensitize our students to some apparent statistical paradoxes.
6. Carry out clinical interviews with our own students to uncover their reasoning processes on probability problems.

Let us consider each of these briefly.

1. From guesses, to experiments, to simulation, to theory.

An experiment with a small-group problem solving, activity-based approach to introductory probability and statistics at the introductory college level was conducted by the author (Shaughnessy, 1977).

The results indicate that students have a better chance of overcoming misconceptions of probability that are due to intuitive heuristics such as representativeness or availability if they are introduced to probability in an experimental activity-based mode rather than in a formal lecture mode. Students can work in small groups, make guesses for the likelihood of events, perform an experiment, organise and analyse the data, and make estimates for the likelihood of events based on their data. More advanced students can perhaps attempt to build a theoretical probability model for the experiment and then compare theoretical predictions to experimental outcomes. The reasons for any discrepancies between experimental results and theoretical predictions, such as sources of bias in the data-gathering process, can be discussed.

An experimental approach to probability may, in fact, be the only way to break the "belief barrier" for those students who do not understand chance at all. Problems which confront misconceptions of statistics can be presented and solved by experiment. Below are several activities that have been used by the author during the first several meetings of an introductory college course on probability and statistics. More activities can be found in previous articles, activities that deal with the effects of sample size on mean, standard deviation, and probabilities of events.

Activity 1: Keep Your Head Up

Before doing this activity, write down your best guess for each of the following: If you flip 6 coins, what is the probability that you will get -

a) 6 heads b) 5 heads c) 4 heads d) 3 heads

Perform this activity in groups of six students:

1. Flip 6 coins. Record the number of heads. Repeat about 50 times. Use your data to answer each of these questions.

    a) What is the probability of getting 6 heads? 5 heads? 4 heads? 3 heads? 2 heads? 1 head? 0 heads?

    b) What is the probability of getting at least 1 head? At least 2 heads?

2. Make a list of all possible outcomes for flipping 6 coins.

    a) Develop a mathematical model to find the theoretical probability for the outcomes of flipping 6 coins.

    b) What is the theoretical probability for getting at least 1 head? At least 2 heads?

    c) What are the assumptions of your mathematical model?

3. Compare the experimental probabilities from part 1 with the theoretical probabilities in part 2. How well do they agree? Make a graph to compare the experimental (observed) probabilities of 50 flips for 0 heads, 1, ..., 6 heads with the theoretical probabilities. Plot the two graphs on the same set of coordinate axes. Where is there close agreement between the two graphs? Where is there not close agreement? Why do you suppose this happens?

4. What assumptions have you made in your experiment when flipping the coins? What suggestions do you have to improve the experiment?

5. List any other comments, questions, observations, or reactions that you might have concerning this activity.

Activity 2: A Pointed Affair

Three different colours (red, gold, and silver in this example) of ordinary household thumbtacks are distributed, one tack to each person.

1. Do this part of the activity by yourself.

   a) First, write down your best guess for the probability that your tack will land upright when dropped.

   b) Devise some uniform way of dropping your tack, and drop it 75 times. Arrange your data in an array with U to indicate upright and D to indicate down.

   c) On the basis of the data you have collected, calculate the probability that the tack will land upright; that it will land down.

2. Do this part of the activity in small groups. Get a person in your group with one each of the three colours of tacks. Use the probabilities calculated in part 1 to list the probabilities that (1) the red tack will land up, (2) the silver tack will land up, and (3) the gold tack will land up. Drop the three tacks together and record the results. Before performing this experiment, write down your best guess for each of the following probabilities:

   a) All the tacks land up

   b) No tack lands up (what's another way to say this?)

   c) At least one tack lands up (what's another way to say this?)

   d) The red tack lands up

   e) Two tacks land up and one lands down

   Drop the three tacks in some uniform way about 60 times. Record your data in an array by listing the reasons as triples; for example, UDD could stand for red tack up, gold tack down, and silver tack down.

   Use the data you have gathered to calculate the experimental probabilities for the events in a-e in part 2. Compare these calculations to your guesses. Any surprises? What assumptions have you made in doing the experiment? What suggestions do you have to improve this experiment?

3. Do this part of the activity in small groups.

   a) Develop a mathematical model to assign theoretical probabilities to the outcomes of this experiment. First list all possible outcomes for the experiment; then devise a way of assigning a probability to each outcome.

   b) Use your data from part 2 to determine experimental probabilities for each of the outcomes listed in 3a. Compare these experimental probabilities with the theoretical probabilities for the outcomes given by your model. Make a graph to compare your experimental probabilities with the theoretical probabilities. How well do the graphs agree?

   c) What assumptions have you made in your mathematical model? Does the model for the tack experiment differ in any way from the model for the coin experiment? If so, how? Is there any similarity between the two experiments?

   d) List any other comments, questions, osbervations, or reactions that you might have to this activity.

Activities such as these are aimed at correcting misconceptions. The tossing of six coins is equivalent to our representativeness problem (1) with the sequence of six children. The tack activity provides experience with unequally likely outcomes so that students begin to realise that if there are N possible outcomes, this does not mean that each outcome has a 1/N chance of occurring. Furthermore, the activities are in two stages. The first stage deals with experimental probability. The second stage encourages students to build their own theoretical models for assigning probabilities.

2. Judgements and Decisions: The confounding role of belief systems vs. the usefulness of statistics in helping to make objective decisions.

   Suppose that you are taking a statistics course which has weekly quizzes with 10 questions each. Your average all year has been 7 correct out of 10. For the next quiz, you study with a friend and get 9 out of 10. Did studying with a friend help?

One possible response to this question is: "Of course it helped, isn't 9 better than 7?" Another possible response is: "Probably not, 9 could have just happened that time since your average of 7 is close to 9". If you ask this question of students in an introductory course in statistics, you will get both the answers above, in addition to a spectrum of "it depends" answers in between. These responses indicate that a belief system is being employed rather than that a calculation has been made. In order to resolve the argument in an objective way, the class needs some data. This is an excellent opportunity to carry out a simulation, say with random numbers, and see just how likely it is to obtain the event 9 out of 10 correct if you have probability of success P(s) = .7 on any one item. (It is not necessary - or advisable in this author's opinion - to do a calculation using the binomial distribution the first time a student sees this type of problem).

Suppose that a consumer's group feels that the gas tanks on a certain model of automobile are likely to explode on impact from the rear. This group has been formed because of the unfortunate death of a friend in just such an accident. The consumer's group has formed a *belief* about a car model. They may even have found other instances of such accidents. Should the car be recalled?

We must attempt to make an objective decision, if possible. Clearly there is some need for more data, and a class in introductory statistics could perhaps be led to suggest some reasonable experiment that can be carried out to determine the safety of the gas tanks. In a problem such as this, the importance of *sampling* in making decisions from statistics can also be brought out. (Would we do our experiments on cars all from one plant? Would we smash *all* the cars?).

If we focus on the usefulness and importance of statistics for making objective decisions, we may be able to temper the effects of belief systems or deterministic approaches to probability in our students.

3. Sensitise students to misuses of statistics.

If we demonstrate the importance of statistics in making objective decisions we are also obliged to show our students how statistics can be used to mislead or to "influence" our belief systems. Books such as *How to Lie with Statistics* (Huff, 1954) should be required reading in all introductory probability and statistics courses. Such books can sensitise our students to misuses of statistics, deliberate or otherwise, and encourage them to analyze probabilistic statements critically.

Students can be asked to keep a log of misuses of statistics that they discover. Articles can be taken from newspapers or periodicals and analyzed for correct or incorrect use of statistics. Students in my classes have ferreted out misleading graphs, inflated percentages, biased samples, insufficient sample size, and verbose statistical quantitative descriptions that had little or no basis in fact.

Many examples of statistics that were based on a biased sample have been analyzed by my students. One found an article on page 1 of the largest-selling newspaper in a major city that mentioned that 95 per cent of the people who were interviewed were against cross-district busing to achieve racial integration in the schools. A continuation of the article (on page 14) mentioned that 70 per cent of the people interviewed did not have children in the schools. Another student criticised the charts published by bookstores and record stores to indicate the top ten books or records for the week. These figures are often based on sales to a particualr class of people with special interests. Yet the list purports to reflect the population as a whole.

4. **Give students challenge problems.**

In order to synthesize the suggestiosn in 1-3 above (i.e., experiments, simulations, confront misconceptions, show usefulness of statistics, sentisize to misuses of statistics) I present challenge problems after several weeks in an introductory course. The students have to turn in a certain number of solutions before the end of the course. They may solve the problems any way they like, by experiment, by simulation, perhaps by a theoretical model if we are far enough in the course or if they do some outside reading. The assignment is called Stimulation with Simulation. Below are several examples of problems from past assignments. The problems have been gathered over the years from several sources or have been created by me. The sources include Mosteller (1962), Mosteller et. al. (1973), Shulte and Choate (1977), and Råde (1977).

STIMULATION BY SIMULATION

The questions in Stimulation by Simulation are somewhat open-ended in nature. In many cases it is necessary to first define clearly what the problem is, and then to adopt a mode of simulation which will enable you to "model" an experiment which may be too cumbersome to actually carry out.

Before doing any of these problems, always write down your best guess first, and only then carry out a simulation of the problem.

A simulation of a probability experiment invovles:

a) modelling an experiment by using some apparatus with known probabilities (i.e. coins, dice, spinners, random numbers).

b) "performing" the experiment many times with your apparatus, thus pooling results of several small groups may help to get a "large enough" sample.

c) gathering, organising, and analysing the data (some statistical skills).

d) calculating experimental probabilities, or other experimental outcomes (i.e. frequencies) from the data.

e) making inferences or drawing conclusions from the experimental outcomes, that is, looking back.

Problems

1. A cereal company has put plastic models of 5 Star Wars characters in boxes of SPACY-O'S. The company puts one character in each box. How many boxes would you expect to have to buy in order to get all five characters? (Write down a guess first).

2. Five friends are sitting around a table at a party discussing the merits of astrology. What is the probability that at least two of them have the same astrological sign? (Assume the signs are equal in duration).

3. Some data were obtained from the records of the Bureau of Meteorology in Perth, Australia. A sunny day was defined as a day on which the officially recorded number of hours of sunshine was 70% or more the maximum possible hours of sunshine for that day of the year. The months April to September in 1973 seemed to be typical months in terms of average number of hours of sunshine. A classification of each day in this period as "sunny" or "dull" (S or D) and recording of the events led to the following:

D followed by D: 79 occasions
D followed by S: 33 occasions
S followed by D: 33 occasions
S followed by S: 37 occasions

Thus the proportion of dull days followed by dull days was 79/112 = .70, and the proportion of sunny days followed by sunny days was 37/70 = .53.

a) What is the "average" length of runs of consecutive dull days?

b) What is the "average" length of runs of consecutive sunny days?

4. Two points (numbers) between 0 and 1 are chosen at random. The points divide the interval from 0 to 1 into three line segmetns. What is the probability that the three segments will form a triangle?

5. About one-half of the babies born are boys. Would there be more days during the calendar year when at least 60% of the babies born were boys:

i) In a small hospital

ii) In a large hospital

iii) Makes no difference

6. Suppose you are playing a game in which you have a 50% chance of winning. Bet a dollar. Each time you lose double our bet. Each time you win, go back to betting a dollar.

Is this a good betting system?

7. Fast Al suggests to Slow Sam that they play the following game. They will each throw either one or two fingers. If Sam throws one and Al throws two, Sam wins $30. If Sam throws two and Al throws one, Sam wins $10. If they throw the same number of fingers, Al wins $20.

a) If each player plays a 50-50 strategy between their two choices, who wins in the long run?

b) What is the best strategy for Slow Sam, i.e. how should he divide his choices?

8. From where he stands, one step toward a cliff would send the drunken man over the cliff. He takes random steps, either towards the cliff or away from it. At any step, the probability that he takes a step toward the cliff is 1/3, and the probability that he steps away from it is 2/3. What is his chance of escaping a fall?

9. 16 teams are entered in a double elimination softball tournament. What is the expected number of games a team will play:

   i) if the team has a 50% chance of winning each game?

   ii) if the team has a 70% chance of winning each game?

10. Several years ago a certain beer company claimed that they "tried their famous High Life taste test on many a beer expert, and no one had yet correctly identified which beer came from a can, which from a bottle, and which from a tap".

    How likely is it that the beer company was telling the truth? Or, perhaps another way to put it, how many beer experts do you suppose the company really tested?

11. Suppose a man faces north and tosses a coin to decide whether to take one step north or one step south. The man does this for twenty-five tries.

    i) On the average, how far is he from the starting point?

    ii) On how many steps is he on the north side of his starting point? On the south side?

    iii) How often does he return to starting point?

12. Three towers are built of 8, 6, and 4 blocks. A tower is chosen at random and a block is removed.

    a) What is the expected number of blocks that will be removed before one of the towers has been "levelled". (i.e., no blocks left?).

    b) What is the probability the 4-block tower will be destroyed?

    c) Can you think of a "real world" situation for which this tower game is a model?

13. (The big-bad-triangle, or three's a crowd). Three duellers, The Good, The Bad, and The Ugly, are to fight a three cornered pistol duel. All know that the Ugly's chance of hitting his target is .30, the Bad's is .50, and the Good (of course!) never misses. They are to fire at their choice of target in succession in the order Ugly, Good, Bad, cyclically until only one man is left unhit. A hit man loses further turns, and is no longer shot at. Ugly has the first shot. What should he do to have the best chance of survival? (Running away doesn't count!).

5. Expose Students to Statistical Paradoxes.

We have already mentioned the importance of sensitizing students to the misuses of statistics and emphasizing the role of statistics in making judgements and decisions. There are times when the way we choose to analyze data, or the underlying assumptions we make in our statistical model, will influence judgements or decisions. It is very important that we expose our students to instances of so-called "paradoxes" so that they remain highly critical of interpretation of data when reading results. If there is more than one possible interpretation for the data, we should encourage our students to examine both interpretations. Several recent articles present instances of so-called paradoxes that are accessible to secondary school students in an introductory probability and statistics class (Reinhardt 1981; Wagner 1981). How, for example, is it possible for baseball player A to have a better batting average than player B both against right-handed pitching and against left-handed pitching, and yet player B gets to start games because she has a better overall average than player A? This example, taken from Reinhardt's article, points out the importance of a weighted average, and the resulting impossible dilemma. Of course when it comes to a decision as to which player should actually start against, say, a left hander, the decision depends on which model the manager chooses to use for the decision: overall average or average against lefties. This is the sort of paradox that can challenge the class to come up with a fairer decision making model.

Another paradox presented in both the Reinhardt and Wagner articles involves the conviction of a couple in a court case based upon six characteristics the witnesses testified were present in the couple at the scene of the crime. The prosecution computed the odds as 1 in 12 million that these characteristics would simultaneously occur, and obtained a conviction based primarily on this calculation. The verdict was subsequently overturned when a mathematician explained that the six characteristics mentioned by the witnesses were far from being clear independent events, and that in fact one could probably find quite a few couples near the scene of the crime with the same six characteristics (Time, 1968). The role of assumptions in our interpretative models is crucial, and we should impress this upon our students.

6. Collect Interview Data from Our Students

As Well et. al. have suggested in their analysis of interviews of college students answering the taxicab problem, there appears to be a great diversity of conceptions and misconceptions of probability and statistics among our naive students. Indeed, the way we tend to think about the subject is not at all the way they (our students) tend to think about the subject. It may be beneficial to both students and teachers if we explored our students' thinking processes in more detail, and reflected on their responses. A numerical estimate for the likelihood of an event, or even a number resulting from some sort of calculation, by itself gives very little indication of how or why that number was chosen. In a recent address, Labinowitz (1982) has suggested that teachers be trained in the art of administering clinical interviews. He has developed a course for preservice and inservice teachers which focuses on the use of audio taped interviews by teachers. The interviews are used to help teachers become aware of student's thinking processes, an exciting approach in an environment which is often all too product oriented. Doing some taped interviews may help narrow the gap between the teacher's projection level and the students' reception level in statistics courses.

As we have seen, some recent research by psychologists and mathematics educators indicates that misconceptions of probability are clearly present in our students. Some misconceptions are intuitive, some are deterministic, and some may even be based upon belief systems. We have offered some possible approaches to the teaching of probability and statistics which may help confront, and hopefully overcome, some of these misconceptions.

References

[1] Bar-Hillel, M. (1980) The Base-Rate Fallacy in Probability Judgements. Acta Psychologica., 44, 211-233.

[2] Edwards, W., Lindman, H. and Savage, L.J. (1963) Bayesian Statistical Analysis for Psychological Research. Psychological Review. 70, 193-242.

[3] Fischbein, E. (1976) Probabilistic Thinking in Children and Adolescents. In: R. Bechauf (ed.) Forschung Zum Prozess des Mathematiklerners. Institut für Didaktik der Mathematik der Universität Bielefeld, 23-42.

[4] Huff, D. (1954) How to Lie with Statistics. London: W.W. Norton.

[5] Kahneman, D. and Tversky, A. (1972) Subjective Probability: A Judgement of Representativeness. Cognitive Psychology, 3, 430-454.

[6] Kahneman, D. and Tversky, A. (1973) Availability: A Heuristic for Judging Frequency and Probability. Cognitive Psychology, 5, 207-232.

[7] Kahneman, D. and Tversky, A. (1974) Judgement Under Uncertainty: Heuristic and Biases. Science, 185, 1124-1131.

[8] Labinowitz, E. Communicating the Results of Clinical Interviews to Teachers Paper presented at the research presessions of the 1982 National Council of Teachers of Mathematics meeting in Toronto. Based upon a writing project at California State University at Northridge.

[9] Mosteller, F. (1962) Fifty Challenging Problems in Probability with Solutions. Reading: Addison-Wesley.

[10] Mosteller, F., Kruskal, W.H., Link, R.F., Peiters, R.S. and Rising, G.R. (1973) Statistics by Example. Reading: Addison-Wesley.

[11] Råde, L. (1977) Take a Chance with Your Calculator: Probability Problems for Programmable Calculators. Forest Grove, Oregon: Dilithium Press.

[12] Reinhardt, H.E. (1981) Some Statistical Paradoxes. In: Teaching Statistics and Probability, the 1981 National Council of Teachers of Mathematics Yearbook, ed. by A. Shulte, 100-109.

[13] Shaughnessy, J.M. (1977) Misconceptions of Probability: an Experiment with a Small-Group, Activity-Based Model Building Approach to Introductory Probability. Educational Studies in Mathematics, 8, 295-316.

[14] Shaughnessy, J.M. (1981) Misconceptions of Probability: From Systematic Errors to Systematic Experiments and Decisions. In: Teaching Statistics and Probability, the 1981 National Council of Teachers of Mathematics Yearbook, Ed. by A. Shulte, 90-100.

[15] Shulte, A.P. and Choate, S. (1977) What are My Chances? Books A and B. Palo Alto, Calif: Creative Publications.

[16] "Trial by Mathematics". (1968) Time, 26, April, p.41.

[17] Tversky, A. and Kahneman, D. (1971) Belief in the Law of Small Numbers. Psychological Bulletin, 76, 105-110.

[18] Tversky, A. and Kahneman, D. (1980) Causal Schemata in Judgement under Uncertainty. In M. Fischbein (Ed.) Progress in Social Psychology. Hillsdale, N.J.: Erlbaum.

[19] Wagner, C.H. (1981) Paradoxes in Sampling. In: Teaching Statistics and Probability, the 1981 National Council of Teachers of Mathematics Yearbook, Ed. by A. Shulte, 162-167.

[20] Well, A.D., Pollatsek, A. and Konold, C. (1982) Probability Estimation and the Use and Neglect of Base-Rate Information. Draft copy of a paper presented at the research presession of the 1982 National Council of Teachers of Mathematics meeting in Toronto. Under revision at the University of Massachusetts, Amherst.

THE ROLE OF EXPLANATION IN TEACHING ELEMENTARY PROBABILITY

J L Myers, University of Massachusetts

Parents, educators, and mathematicians have been involved in recurrent debates about the relative merits of understanding mathematical concepts and rote memorizing computational formulas. In principle, such conflicts should be resolvable through research involving manipulation of the degree to which students understand mathematical material and observation of their performance on tests of mathematical knowledge. In actuality, we encounter difficulties because we lack both a precise definition of mathematical understanding and a specification of the processes such understanding might affect.

We have begun a research program to investigate the role of understanding in learning elementary probability by university undergraduates. Although we still lack a precise definition of understanding, we have identified three dimensions which have provided the basis for constructing short texts. The first such dimension, external connectedness, corresponds to the degree to which a text explicitly relates the concepts and formulas to be learned to the real-world knowledge the student presumably brings to the situation. The second dimension might be termed "internal connectedness", the degree to which relations among formulas and concepts are explicitly developed. For example, an equation to be learned may, or may not, be presented as a special case of a formula presented earlier in the text. The third dimension is explanation, the degree to which a rationale is provided for a particular formula. This refers not to derivations, but to intuitive justifications for procedures taught. For example one of our texts uses branch diagrams to explain why the probability of the joint occurrence of two independent events is a product of the marginal probabilities.

We will describe the texts constructed within this framework more fully in a later section. First, however, we wish to consider certain processes which should be involved in solving probability problems, and the ways in which these may be affected by the degree of understanding imparted by the text. We assume three stages: categorization of the problem, retrieval of the formula appropriate to that category, and translation of the problem - substitution of values from the statement of the problem into the retrieved formula. In general, understanding should aid both the categorization and translation of problems. Understanding may aid or hinder retrieval of formulas from memory. On the one hand, the linkage of ideas and formulas make the formulas easier to learn and recall; on the other hand, the additional explanatory and integrative material could detract from time and effort that should be spent learning the equations.

From our perspective, it makes little sense to examine total test scores as a function of the degree of understanding incorporated into the text used for instruction. Which text is better may well depend upon which processes are most important in the solution to a particular problem. For example, assume that texts which emphasize rote learning of formulas at the cost of de-emphasizing explanation do provide an advantage in memorizing formulas. Then, problems in which categorization and translation are not difficult may be more easily solved by students who have less understanding of the material. On the other hand, assume that a problem provides no key words which immediately signal the problem's category, or contains considerable irrelevant information which make categorization or translation difficult; a text which has emphasized understanding may provide a definite advantage here.

In view of these considerations, we have attempted a more molecular analysis of test performance than is usually the case. For each of several text conditions, we have examined variations in performance as a function of the type of problem. At a rather gross level, we have formula problems, which merely presented certain values (e.g. "P(A) = .6, P(B) = .4") and necessary conditions ("A and B are independent events"), and required the subject to set up the correct results (e.g. "What is P(A and B)?" requires .6 x .4 as the answer); and story problems which were considerably less transparent. Within the set of story problems, performance has been related to such variables as the presence or absence of key words (e.g. "and" implies the multiplication of probabilities), the presence of irrelevant values, and translation requirements (e.g. whether it was necessary to subtract a stated probability from one in order to solve).

Our perspective has also led us to require subjects to report their answers as operations on numbers; for example, a correct answer would be .4 x .2, not .08. This approach also provided more detailed knowledge of the types of errors made.

To summarize, we view understanding of mathematical material as a function of (1) connections of text concepts and formulas to real-world referrents; (2) integration of concepts and formulas within the text; and (3) explanation of formulas. This view has provided a basis for constructing three written treatments of elementary probability, presumably varying in the degree to which they convey understanding. Our view of the processes involved in solving problems has led us to use both formula and story problems, and to emphasize analyses of error protocols.

METHOD

Subjects

Forty-eight volunteers from an Introductory Psychology course at the University of Massachusetts were randomly assigned to three text conditions. None had previous exposure to probability or statistics.

Texts

There were three texts, each judged to be representative of actual chapters on probability in various introductory statistics textbooks. All contained the six formulas shown in Table 1, with one numerical example illustrating the application of each formula. All three texts began with a brief introduction indicating that the text contained a series of examples of probability calculations, and that each example would be followed by a formula which could be used to solve problems of that type. All three texts then contained a section labelled "What are the chances?" which introduced the concept of probability and its limiting values. This section noted that an action such as tossing a coin or observing the day's weather could have several possible outcomes, that a probability value could be assigned to these outcomes, and that these values ranged from zero to one.

TABLE 1

Formulas Presented in the Three Texts

(1) $P(A \text{ or } B \text{ or } C) = P(A) + P(B) + P(C)$
(A, B and C are mutually exclusive)

(2) $P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$
(general case)

(3) $P(A \text{ and } B \text{ and } C) = P(A) \times P(B) \times P(C)$
(A, B and C are independent)

(4) $P(A, B, B, B, A) = P(A) \times P(B) \times P(B) \times P(B) \times P(A)$

(5) $P(A \text{ given } B) = P(A \text{ and } B)/P(B)$

(6) $P(A \text{ and } B) = P(A \text{ given } B) \times P(B)$

Despite these similarities, the texts differed in several ways. A *high explanatory* text defined probability as a relative frequency of events, and examples emphasized the relation between answers and counts of the numbers of ways that various events could occur. A *standard* text treated probability as a measure with certain well-defined properties (e.g. limits of one and zero) and used familiar real-world examples (e.g. the probability of rain tomorrow), but did not explain probability in terms of counting processes. A low-explanatory text was more abstract than either of the first two; aside from an initial reference to dice-throwing, examples were stated in terms of such quantities as P(A) and P(B).

The texts also varied with respect to the degree of linkage among the equations presented. Only the high-explanatory text showed that one equation was a special case of another, or that one equation could be derived from another. For example, this text pointed out that the probability of the union of two mutually exclusive events (Equation 1), was a special case of the probability of the union of any two events (Equation 2).

Finally, the texts varied in the degree of explanation provided for each equation. In the high-explanatory text, Venn and branch diagrams were used to represent the information in the examples, and the equations applied in these examples were explained in terms of these diagrams. Neither of the other texts provided such explanations. For example, although all three texts stated that the probability of the joint occurrence of two independent events was the product of their probabilities and provided an example of this, only the high-explanatory text attempted to demonstrate why probabilities were multiplied in such cases.

In summary, all texts contained an example followed by the equation, as well as sufficient definitions of terms such as mutual exclusivity and independence. In the high-explanatory text, the relation of probability to counting was emphasized, pictorial aids were used, and the equation was presented as a natural development ensuing from the example.

Performance Tests

Two tests were constructed, each containign two sets of six story problems and two sets of six formula problems. Within each set, there was one problem testing each of the six equations in the text. The order of sets within a test was counterbalanced over subjects so that there were four sequences, with the constraint that story and formula sets would alternate. Half of the subjects received one test first and the remaining subjects received the other test first.

Formula and story problems differed in the degree to which the given information was explicit in the question. For example, a formula problem testing knowledge of Equation 1 would state "P(A) = .4 and P(B) = .3. If A and B are incompatible, what is the value of P(A or B)?". In contrast, a story problem testing the same knowledge would read as follows: "The probability is .20 that Jimmy will win a race and .30 that he will finish second. What is the probability that he will win or finish second?".

Procedure

Subjects were tested individually. They were instructed to read the text booklets silently and at their own rate, trying to understand the explanations and examples. The standard and low-explanatory groups, whose texts were four pages in length, were told that they would have 15 minutes to finish; the high-explanatory group, whose text was 14 pages in length, was given 25 minutes. These limits, based on pilot subjects, were imposed to discourage any individual from either quickly scanning the texts, or taking an inordinately long time. All subjects finished reading their texts in the given time period for their group. Subjects in the low-explanatory and standard groups tended to rate the study time more favourably than subjects in the high-explanatory text condition, but this difference was not significant.

At the completion of the study period, subjects were given two minutes to study a list of 15 words unrelated to probability and were then required to write down as many words as they could recall on a blank sheet of paper.

Upon completion of the interpolated task, subjects were given the test. There was one problem on a page with space to do any work which the subject felt might help in the solution. Subjects were given as much time as they required to complete the test; most required about 30 minutes. Upon completing the problems, subjects were asked to return 48 hours later. On the second test day, subjects were given the second form of the test. The order of presentation of the test forms was counterbalanced.

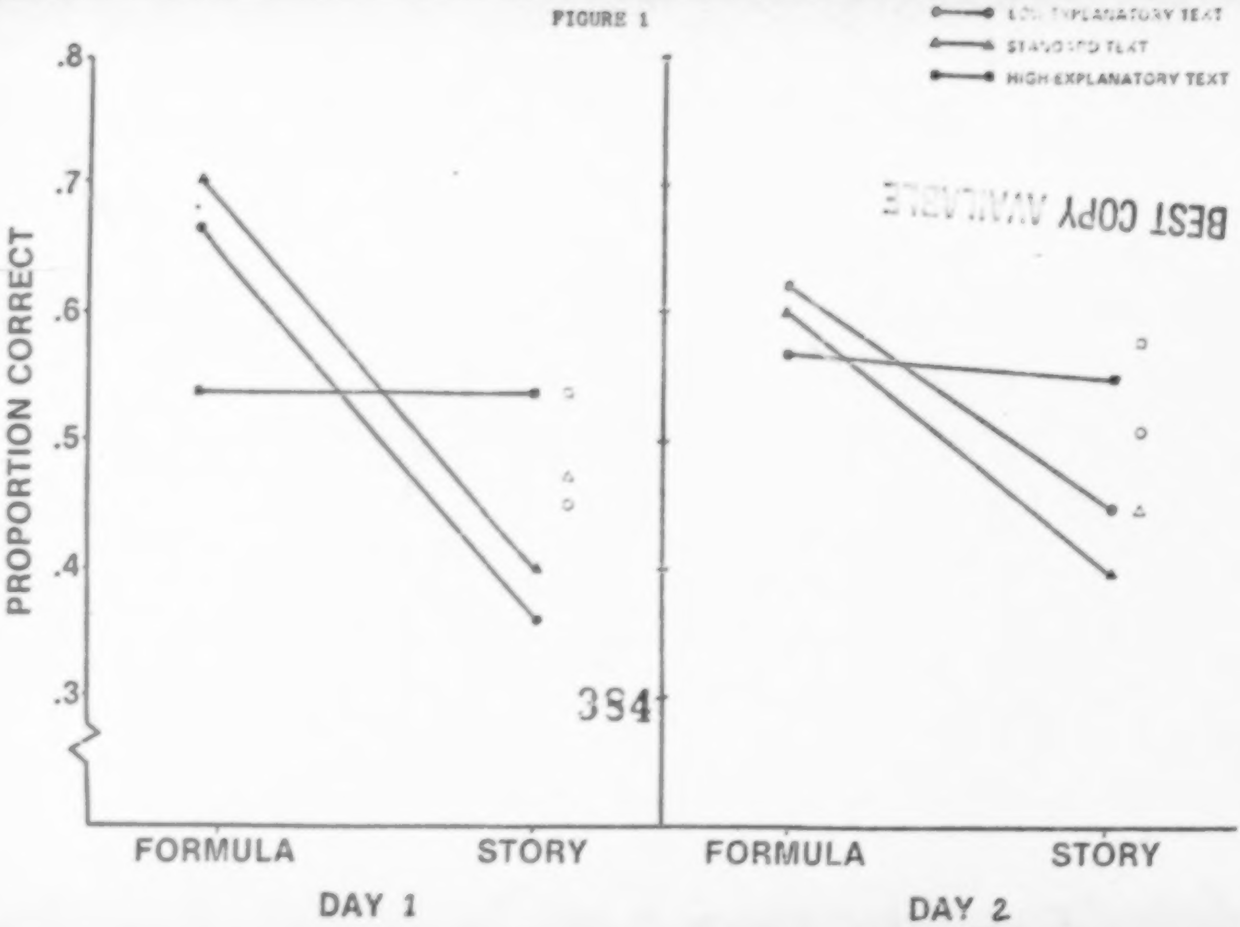
RESULTS AND DISCUSSION

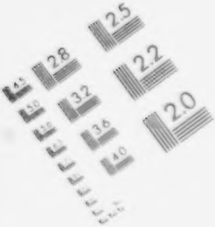
Performance test data were scored according to both a strict and a lenient criterion. The strict criterion required that the subject set up the answer correctly in all respects; the lenient criterion gave credit if the subject used the correct equation but, for example, inserted the wrong values. This distinction applies only to performances on story problems; strict and lenient scores were the same for formula problems.

A Summary of Test Data

Figure 1 summarizes the data. The points on the lines are for the strict criterion; the disconnected open symbols are for the lenient criterion. There are two important observations to be made about the means. First, considering the strict-criterion means, the relative performances of the three text conditions strongly depend upon the type of problem ($p < .001$). The two non-explanatory groups appear to know the formulas better than the high-explanatory group but they have more difficulty in applying that knowledge to story problems.

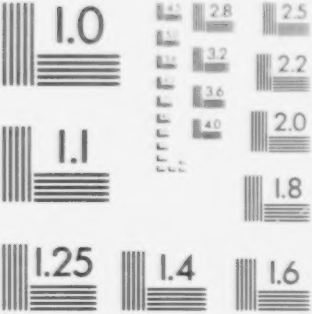
FIGURE 1


LOW EXPLANATORY TEXT
STANDARD TEXT
HIGH EXPLANATORY TEXT
PROPORTION CORRECT
.8
.7
.6
.5
.4
.3
FORMULA
STORY
DAY 1
FORMULA
STORY
DAY 2



PG

100 mm



ABCDEFGHIJKLMNOPQRSTUVWXYZ
abcdefghijklmnopqrstuvwxyz1234567890

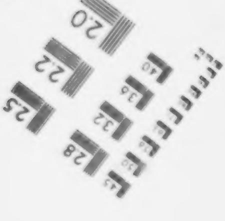
ABCDEFGHIJKLMNOPQRSTUVWXYZ
abcdefghijklmnopqrstuvwxyz
1234567890

1.0 mm

1.5 mm

2.0 mm

A5



The second point to note in Figure 1 is that strict-criterion scores are significantly lower than lenient-criterion scores for the two non-explanatory groups ($p < .001$). Apparently, subjects in these two groups may correctly categorize the problem and recall the appropriate formula but are not able to translate the values in the problem statement. In contrast, when high-explanatory subjects know what operations are to be performed to solve a problem, they usually perform them on the correct set of values; there is little difference btweeen their strict and lenient-criterion scores.

A More Detailed Analysis of Individual Problems

It is impractical to examine the numbers and types of errors made to each of the 24 story problems for each of the three text conditions. Therefore, we will focus on those problems testing knowledge of Equations 1, 3, and 4: the problems within these sets reveal the widest variation in proportions of correct respones. In considering these problems, we ask: why is there marked variation in proportion correct to problems requiring the same equation for solution? And why is this pattern of variation affected by text condition?

The low-explanatory and standard groups performed in similar fashion on all problems. Therefore, in the interests of simplifying the presetnation, we have treated them as a single condition which we have labelled "non-explanatory." We will refer to the high-explanatory group as "explanatory".

Equation 1. Performances on these four problems indicate that subjects in the non-explanatory condition are sensitive to aspects of the problem statement which have little effect upon the performance of subjects in the explanatory condition. Proportions of correct responses in the non-explanatory group varied greatly across the four problems testing knowledge of Equation 1: (a) .375 (b) .4375 (c) .625 and (d) .7188. In contrast, the performances of those in the explanatory group are not only much better but also quite stable: (a) .9375 (b) .75 (c) .8125 and (d) .8125.

Examining the answers to (a) and (b) more closely, we find clear evidence of failures to categorize or translate these problems correctly, but only in the non-explanatory condition. Consider problem (b):

> "In the National Women's Ping Pong League, Los Angeles has a .3 chance of winning, San Francisco has a .2 chance, San Diego has a .1 chance, New York has a .1 chance, and Philadelphia has a .3 chance. What is the probability that a west coast team will win?"

Note that this problem contains irrelevant information; the stated probabilities for New York and Philadelphia are unnecessary. Subjects in the non-explanatory condition attempted to use these probabilities; five subjects subtracted .1 + .3 from the corect sum (or from the product of the west coast probabilities), four others multiplied all five probabilities, and two others divided by .1 + .3. Only one subject in the explanatory condition attempted to incorporate the irrelevant values into his answer. Apparently, subjects who do not understand the rationale for using the equations attempt to use any information presented to them. In this example, the subjects cited misclassified the problem, using Equations 2, 4 or 5 in an attempt to incorporate all the values stated.

Problem (b) also indicates that subjects in the non-explanatory condition have difficulty when there is no key word to provide a basis for classification. Eight of the 32 subjects in this condition used Equation 3 rather than Equation 1, multiplying probabilities instead of adding them. It is revealing that the only other of the four problems in which this type of error was frequent (nine of 12 subjects who made errors) was (c); (b) and (c) are the only two problems in this set in which the problem does not specifically request the probability of something or something. It is also important to note that this misclassification error occurred only once for problem (b) and once for problem (c) in the explanatory condition.

Even when subjects in the non-explanatory condition used the correct equation, they may have had difficulty translating the story; that is, correctly inserting values from the story into the equation. This was a major source of difficulty for the non-explanatory group in problem (a). The problem presents frequencies of different coloured marbles in a jar, and then asks for the probability of drawing a red or white marble. Five subjects misclassified the problem, attempting to incorporate frequencies of orange and blue marbles into their answer. Most of the errors, however, reveal a lack of understanding of the relation between frequency and probability. Subjects often failed to divide the sum of the numbers of red and white marbles by the total number of marbles, or divided by 100. Note that this problem was the most difficult for the non-explanatory group but the easiest for the explanatory group who had the relation between frequency and probability explained in their text.

The performance of the non-explanatory group on problem (d) provides further evidence of the validity of our analysis of classification and translation errors. This problem evoked the best performance from the non-explanatory group. Not surprisingly, in view of our analysis, it states exactly two values, includes the key word "or", and requires no translation of the stated values.

Equation 3. The explanatory group did not perform as well on these problems as on those testing knowledge of Equation 1. We suspect that the rationale for multiplication of probabilities is more difficult to understand than that for adding probabilities.

In one respect, performances on this problem set were similar to those on the first problem set: proportion correct was fairly stable over problems for the explanatory group and quite variable for the non-explanatory group. For the explanatory group, these proportions were (a) .5 (b) .5625 (c) .625 and (d) .5625; for the non-explanatory group, the proportions were (a) .2813 (b) .5 (c) .625 and (d) .8438. As with the first problem set, it appears that subjects in the non-explanatory condition are more sensitive to the statement of the problem than are subjects in the explanatory condition.

Much of the variability exhibited by the non-explanatory group can again be traced to variations in the problems with respect to irrelevant information and translation requirements. Problem (b) presented values that were redundant (in essence, presenting P(A), P(not-A), P(B), P(not-B)). Four subjects in the non-explanatory group used the wrong equation in an attempt to incorporate all four values in their answer (e.g. ".2 x .4 x .6 x .8") and one made what we would characterize as a translation error, multiplying .2 by .4. In contrast, no subject in the explanatory group used numbers other than .2 and .6 in their answers.

The answers to problem (a) indicate a serious translation difficulty. The problem states P(A) and P(not-B), and asks for P(A and B). Of 23 subjects in the non-explanatory condition who erred in response to this problem, 11 multiplied P(A) by P(B); only one subject in the explanatory condition did this.

In contrast to performances on problem (a), the non-explanatory group did very well in answering problem (d). This states:

> ".7 of the members of the state legislature favor a particular bill. .6 of the members are Democrats. If party and vote are independent, what is the probability that the next legislator you meet is a Democrat and votes for the bill?"

We suggest that this problem was very easy for the non-explanatory group because they have memorized the rule: "Whenever the probabilities of something and something else are required, multiply their individual probabilities". In problem (d), the word "and" is present and there are only two possible values to multiply. Fortunately, they are the right two values. The same circumstances hold for problem (a) except that the values given are not right; one value must be subtracted from 1.0 before multiplying.

Equation 4. This equation is the general case of Equation 3 and, therefore, it would seem that subjects should score about the same on this problem set as on the earlier one. That was the case for subjects in the explanatory group;

in the non-explanatory group, however, the average proportion correct was .22 lower for the problems testing knowledge of Equation 4 than for those testing knowledge of Equation 3. For the explanatory condition, the proportions correct were, as in the other problem sets, relatively stable: (a) .5 (b) .4375 (c) .625 and (d) .5625. Also, as before, the non-explanatory group exhibits greater variability of performance over the four problems: the relevant proportions were (a) .1563 (b) .1875 (c) .3438 and (d) .625.

That the non-explanatory had greater difficulty with these four problems than with the third problem set may reflect two factors. First, these problems typically omit explicit use of the key word "and"; this may cause classification problems. Second, the probability of the complementary event is often not explicitly stated; thus, there may be translation difficulties.

Both conditions exhibit about 16% classification errors as indicated by use of operations other than multiplication. More striking, at least in the non-explanatory condition, were errors of translation indicated by using the stated probability when its complement was appropriate. Averaging over problems (a), (b) and (c), 11.67 subjects in the non-explanatory group made such errors. Only 2.33 subjects did so in the explanatory group.

The non-explanatory group performed reasonably well on problem (d) which states:

> "A hockey team wins with probability .5, loses with probability of .3, and ties with probability of .2. What is the probability that in the next four games, the team wins, then loses, then wins, then loses?"

The statement of this problem more closely parallels the statement of Equation 4 than does that of the other problems in this set, thus reducing the likelihood of misclassification. Furthermore, all the required probabilities are stated; translation difficulties are presumably minimal.

CONCLUSIONS

Very different patterns of knowledge appear to be present in subjects in the non-explanatory (standard and low-explanatory texts) and explanatory conditions. Subjects in the two non-explanatory groups performed considerably less well on story than on formula problems, and often used the correct formula for a problem (that is, met the lenient criterion) but failed to solve it. A closer look at answers to story problems revealed that subjects in the non-explanatory conditions often required the explicit presence of key words which unambiguously pointed to certain operations, tended to misclassify problems in the presence of irrelevant or redundant information, and made many errors when the values in the story required modification before insertion into the formula. In contrast, subjects in the highly-explanatory condition performed equally well on story and formula problems, tended to solve whenever they showed evidence of knowing the appropriate formula, and were considerably less hindered by absence of key words, the presence of irrelevant information, and the need to translate values in the story. The situation may be summarized by noting that, although all our subjects were novices, those in the high-explanatory condition seem to share more of the characteristics of experts than do those in the other two conditions.

The results of more recent studies lead us to believe that the solution processes in the explanatory and non-explanatory conditions are qualitatively different. In one study, subjects in a high-explanatory condition had only 60% correct recall of equations (as opposed to 90% for a standard group), but were able to perform correctly on 53% of story problems. This result suggests to us (as does the equivalence of formula and story scores in the study just detailed) that subjects in the high-explanatory condition often may not retrieve the formula at all. Another study in which subjects thought aloud while solving supports this conjecture; after studying a high-explanatory text, students tended to attempt to construct solutions, after using the diagrammatic aids provided by the text. Rather than recalling formulas, they attempted to recall examples from the text which they felt were similar to the problem at hand, and to use these as a model for solution.

Much remains to be done. High on the agenda must be the use of other measures of comprehension - both problems demanding more complex processing and transfer to new materials such as the Binomial and Bayes' theorems. We suspect that such research will provide further evidence of the role understanding can play in mathematical problem-solving.

Acknowledgements

This research was supported by a grant from the National Institute of Education, NIE G80-0126, and by a grant from the National Science Foundation, BNS 80-01181. Randall Hansen, Joan McCann, and Rikki Robson made major contributions to all phases of the research reported in this paper.

PROBABILITIES: LEARNING BY DOING OR LEARNING BY IMITATING

A Abele, Pädagogische Hochschule, Heidelberg

INTRODUCTION

Learning mathematical terms like 'frequency', 'random event', 'probability' and the like is closely connected with the means of illustration that you - the teacher or the pupil - choose. In solving descriptive mathematical problems, school children can be grouped into two different types: those who prefer graphs and graphical procedures like situational outlines or diagrams and the like; and those who prefer to choose a verbal form of expression and who like to work with symbolic means.

The learning situations for the first type of learners have to satisfy two conditions:

(1) The teacher has to organize his instruction as the external condition for the learning approach of the pupil in such a way that the pupil is able to develop his creative abilities in looking for 'roundabout solutions'.

(2) This is particularly successful when graphs can be used as an iconographical means that can be actively employed by the pupil. Even incorrect solutions may support the recognition process of the pupil, because a comparison of the graphical trial solution with the situation that is to be depicted that is undertaken by the pupil himself gives him much more insight than a correcting direct hint of the teacher.

INVENTION OF GRAPHICAL DESCRIPTIONS

The application of graphical means of solving an existing problem is, in a school situation, almost always enhanced by the teacher. One specific graph that had proved to be particularly apt for solving the pertaining problem is taught to the pupils as the standard procedure. As a rule, the pupil does not take part in the development process of this procedure, instead, it has already been 'invented' and is merely accepted by the pupil. If this attitude is to be corrected, one has to know how graphical solutions are arrived at via preliminary or even incorrect intermediary steps or phases.

The first questions are:

- What do children of different ages do in specific situations when the teacher wants them to invent a graphical description of a random process?
- Is it possible to bridge the gap between the pupil's proposal and the standard solution which is in the mind of the teacher?

We look at two typical examples:

390

1st example: Repetitive throwing of a coin (grade 7/8)
2nd example: Passing four roads by one/two/three cars by chance (grade 8).

First Example: Repetitive Throwing of a Coin

We throw a coin twice, three times, four times etc. Which combinations of head and tail could there be in such a series of trials? The solutions of the pupils are very informative. In confronting the pupils' proposals with the standard solution which is in the teacher's mind we recognize the gap between these two ways of solving the problem.

As a rule we solve the problem

(1) with graphical means using the tree diagram

The coin was thrown

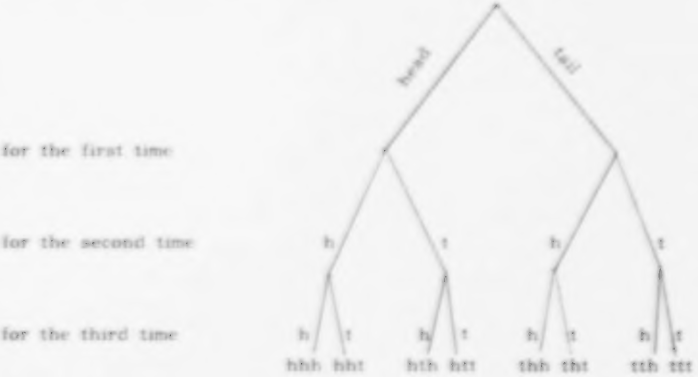


FIGURE 1

(2) or by calculating all combinations (variations with repetition), which are possible through two letters (or colours) on three places.

It has to be mentioned that the children had not been made familiar with either of the two methods. Some of the pupils declared they had not learned how to solve problems like this and waited for some help from their teacher. Others, however, set about solving the problem. In doing so one of the pupils proposed the following solution. He selected the iconographical level and came up with the following outline.

The coin was thrown

for the first time

for the second time

for the third time

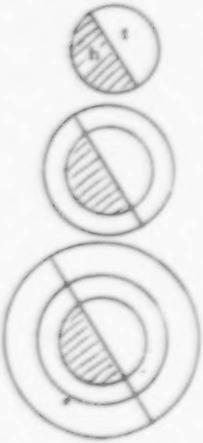


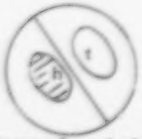
FIGURE 2

This illustration could only be applied to the situation when the coin was thrown for the first time. It is obvious that it could not be generalized.

Another pupil presented the following very remarkable solution.

The coin was thrown

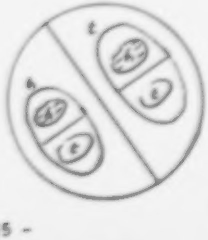
for the first time



He is using something similar to a "Venn diagram" and therefore in his thinking he has to put the elements of the set - the two events, which are represented by the two coloured zones - inside the boundary line of the Venn diagram.

What may the pupil do in applying his first proposal?

for the second time



for the third time

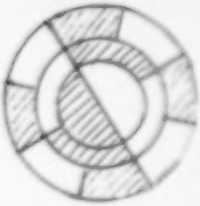


FIGURE 3

This is a correct solution and an illustration that could be generalized - found by the pupil. For that reason it must not be ignored even when we think it would be necessary for the children to learn the concept of decision-tree and its application. The approach taken by the pupil leads to the following consequence: From a methodological point of view it may often be wrong to have the pupils learn a standard procedure introduced by the teacher with the assumption that they will afterwards learn other procedures or inconographical representations. On the contrary we see the possibility of organizing the lesson in such a way that the pupil is able to develop his creative abilities in looking for self-made illustrations.

Second Example: Passing Four Roads by One, or Two, or Three... Cars by Chance

Four children of 8 grade are working with a map. Two villages A and B are connected by four roads.

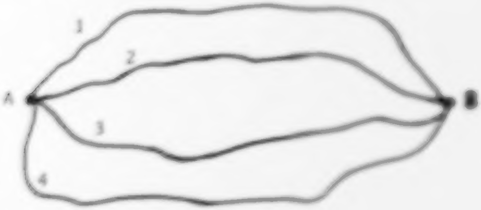


FIGURE 4

First problem: Two cars (red respectively yellow) run from A to B. How many possibilities have they to choose a road, if they run independently and by chance? The children try to solve the problem with the understanding that the yellow car takes another road than the red car. They presented their solution using two different illustrations.

first proposal

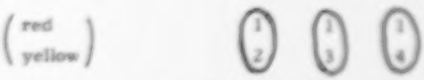


FIGURE 5

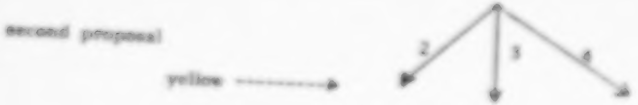


FIGURE 6

In every case the solution is not complete. The children do not notice the other possibilities. In these 9 cases the red car does not pass road one but road two or three, or four:

- red road 2 and yellow road 1 or road 3 or road 4
- red road 3 and yellow road 1 or road 2 or road 4
- red road 4 and yellow road 1 or road 2 or road 3

Second problem: How many possibilities have the two cars to choose one of the four roads by chance if the yellow car does not take necessarily a different road from the red car? The pupils suggest a new illustration.

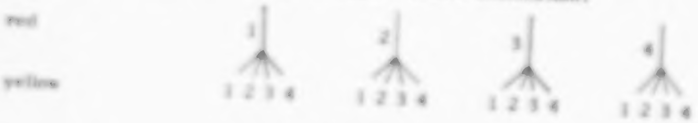


FIGURE 7

I want to accentuate two points:

- The branches relating to the red car do not start from one point
- The children do not use the illustration of ordered pairs $\binom{1}{2}$, as they did above.

Third problem: One pupil makes the following suppositions:

- The red car and the yellow car take different roads.
- Each of the two cars gets two of the four roads to choose one of them, for instance: red gets the roads 1 and 2, yellow the roads 3 and 4.

For this part of the problem the solution is illustrated by the pupil twice: "decision tree" as above

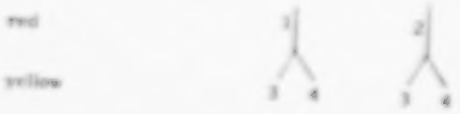


FIGURE 8

ordered pairs in the sequence (1,3) (2,3) (1,4) (2,4)

But nobody sees the other 20 possible pairs as further solutions. This is a typical phenomenon which is caused by the specific difficulty of probabilistic problems: For the pupils nothing is a real situation, which could have been mathematicized, all is hypothetical. If they pick out the four cases, which can be realized at the same time, they have created their reality in their mind and the fictive totality of the 24 cases has disappeared.

Fourth problem: Three cars run from A to B. How many possibilities have they to choose a road, if they run independently and by chance?

First the children constructed the following illustration as a continuation of the figure they had used before.

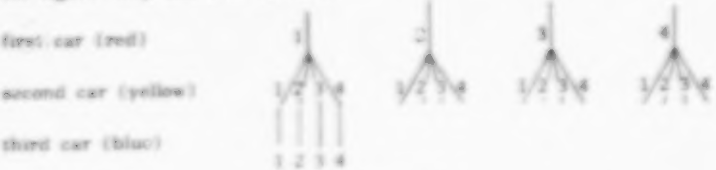
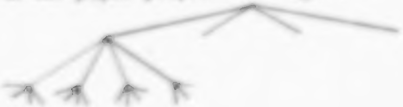


FIGURE 9

But this illustration is not suitable for a propitious description of this situation. After this first step one of the pupils proposes this figure



which is in use everywhere.

FIGURE 10

Fifth problem: What about 4,5,6 etc. cars and the four roads - how many possibilities? This is now very easy. One of the pupils mentioned: "Always this is quadrupling". They no longer used an iconographic representation because now they had the general structure in their imagination.

PROPOSALS OF LEARNING ACTIVITIES IN THE "HAUPTSCHULE" (SECONDARY SCHOOL FOR SLOW LEARNERS)

Taking the observations above mentioned and the like as a basis we now want to offer some proposals for starting activities in the learning of probabilities in the "Hauptschule", which is a special secondary school for children on a very poor and low level. The proposals are both problem oriented, the first referring to some combinational problems which are very challenging for the pupils, the second starting with a well-known game whose analysis leads to the concept of probability.

Combinational problems

- Initial problem: Lotto, 6 numbers are drawn out of 49 numbers.
- Second problem (reduced): Lotto, 2 numbers out of 5.

- Third problem: Lotto, 2 numbers out of 6.
- Fourth problem: Lotto, 3 numbers out of 6.
- Fifth problem: Permutations
- Generalization
- Combinational basic problems

The concept of probability

- Initial game: "Mensch ärgere dich nicht" (game at dice).
- Analysis of the game by isolating single situations
- Probability, first definition.
- Decision tree and path diagram: Throwing a die two times, three times etc.
- Game at dice using four dice.

HOW DO PUPILS OF A SECOND GRADE HANDLE COMBINATIONAL PROBLEMS? DESIGNING COLOUR PATTERNS AND COMBINATORIAL COUNTING

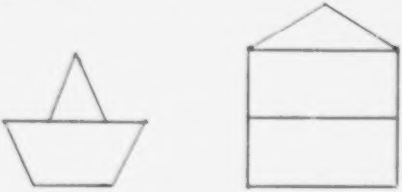
As we all know, group instructional methods promote the learning of mathematics in situations of exercising. If - within a well-defined area - certain techniques could be trained and retained, group specific phases of work have a stimulating effect on the performance and the social behaviour of the pupil. In these cases group situations promote mathematics-related discussion and goal-oriented actions of the children. Therefore it is supposed that group instruction in mathematics namely in probabilities also has positive effects for the identification of new insight. So we expect - according to some preliminary studies - the following from the pupils' work:

- The pupils can - without major assistance - dissect the problem into parts of problems and can thus work on these step by step before they reach a solution.
- The working material and the task assignment will enhance in the pupils a problem-specific willingness to discuss which is the basis for successful group work.
- The group situation will create possibilities to formulate solution hypotheses and to discuss and review conceptual suggestions on the basis of arrangements of the materials which can also be quickly discarded. Drawing, which takes time, and written notes, which bear the inherent and well-known difficulties of subsequent correction, are not necessary.

Didactical Analysis

Four children of one group of the second grade receive paper shapes in different colours with which they are supposed to lay out houses and ships. After a first, relatively free working phase, the teacher suggests to the pupils to lay out ships from two paper shapes and houses from three paper shapes according to Figure 11 and to try to find as many colour combinations as possible.

FIGURE 11



Ship House

Through these instructions the mathematical problem is put into a concrete setting and is thus accessible for the pupils on the inherent active expressive level. The aim of the project is the gradual systemization of all colour variations (variations with repetition), which are possible through three colours in two places and through two colours in three places. The survey solution - in this form not really the direct aim of the pupils' work - is easily seen with the aid of a decision tree.

There are a total of nine possible colour combinations according to this rule.

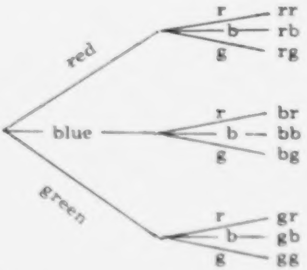


FIGURE 12 - Decision Tree A: Ships

There are a total of eight possible colour combinations according to this rule.

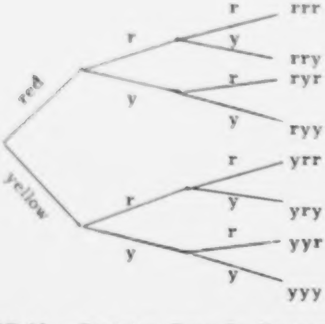


FIGURE 13 - Decision Tree B: Houses

Sequence of the group work

During the childrens' work one can differentiate clearly among five work phases:

First phase: Free construction

Second phase: Construction according to a rule - preliminary solution

Third phase: Review of the preliminary solution

Fourth phase: Finding individual aspects - beginning systemization

Fifth phase: Reaching a total solution through combination of the individual aspects - complete conscious systemization

Thus, without formulating it so specifically, the total problem is divided by the children into the following assignment parts:

1st part assignment: We build a great number of houses or ships (any kind)

2nd part assignment: We build a great number of houses or ships (specific)

3rd part assignment: We look for duplicates and try to fit the rest of the paper shapes in also

4th part assignment: We sort out all houses or ships

5th part assignment: We draw or note the results in a distinct form

The second phase was introduced by the teacher with the request that the pupils limit themselves to specific types of ships and houses; the fourth phase through the request to see "how many houses and ships are there". On the other hand, the transition from the second to the third and from the fourth to the fifth phase are fluent; they came about without the help of the teacher through discussion among the pupils themselves.

A careful analysis shows some remarkable details in the attitudes of the pupils.

First phase: Free construction

The pupils strive to construct a great variety of forms. They design two types of ship and four types of house (see Figures 14 and 15) which they offer in the various colour combinations.

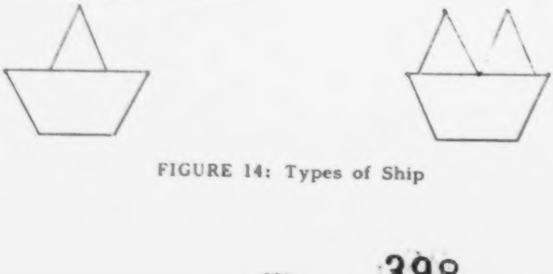


FIGURE 14: Types of Ship

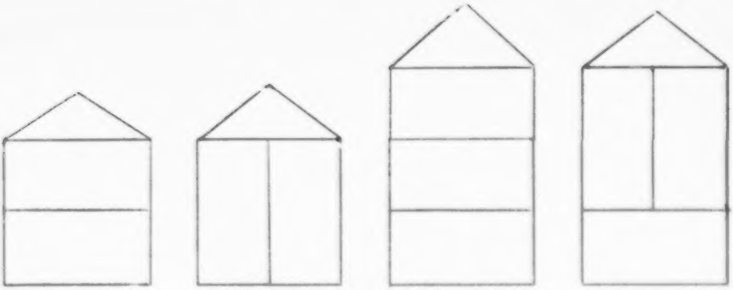
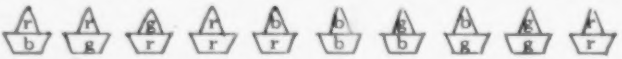


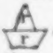
FIGURE 15: Types of House

For the variations aesthetic colour arrangements were above all decisive ("red and blue go well together"), the quantitative aspect, to find a great number of different colour combinations hardly played a role in this phase. The teacher intervenes when nine ships and nine houses are built and introduces the second phase with the additional rule.

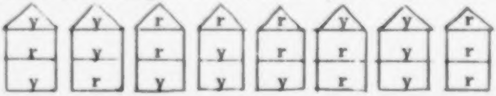
Second phase: Construction according to a rule
The pupils work in two directions, some rebuild the existing houses and ships to fit the requested forms, others construct new ships from the existing paper shapes. This work division without mutual consultation leads to the fact that some ships and houses occur twice. When this is noticed ("These are double, that we already have up there, now we have to put in a different colour"), the pupils begin to control the results that they have so far produced.

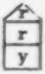
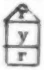
Third phase: Review of the preliminary solution
In order to find out whether "everything is all right", the children adopt two strategies: they sort out grossly separating ships from houses which, at the beginning, were mixed together and search for doubles. Moreover, the children try to build more figures with the rest of the shapes ("Maybe now we get it all done"). Toward the end of this control and supplementary phase ten ships and eight houses are laid out on the table:
Ships of two colours consisting of the colours red (r), blue (b) and green (g):

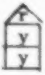
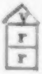


The combination , which occurred for the second time was obviously not seen by the pupils.

Houses of two colours consisting of red (r) and yellow (y):



Here again, it was not seen that the combination  occurs twice. The combination  is, however, missing.

In this control activity the pupils were invariably led by aesthetic viewpoints or the expectation of a certain regularity ("Look here, this is red-red, yellow-yellow", or: "yellow-red, red-yellow, those two there..." What was meant were the colour combinations  and ).

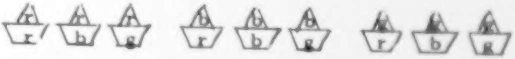
These beginnings of a systematization can be recognized more in the next phase, which is introduced through the teacher's question asking for the number of the figures.

Fourth phase: Finding individual aspects

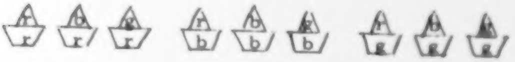
The question of how many houses or ships is initially quickly answered. When a more differentiated questioning occurs, the pupils follow different classification viewpoints.

Classification of ships

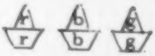
a) Grouping according to sails and boat hulls of the same colour. This method follows the taxonomy of the decision tree A (Figure 12) and offers the following three groups



or



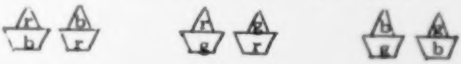
b) Groupings according to one-colour or multi-colour. You get three one-colour ships



and 6 multi-colour ships.

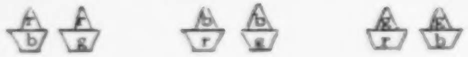
c) First continuation of b)

Subdivision of the multi-colour ships alternating in pairs



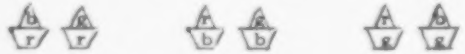
d) **Second continuation of b)**

Subdivision of the multi-colour ships in pairs according to sails



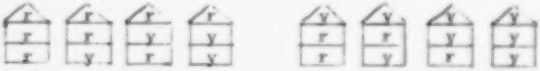
e) Third continuation of b)

Subdivision of the multi-colour ships in pairs according to hulls

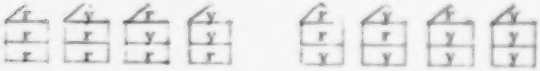


Classification of houses

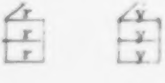
f) Grouping according to roofs and floors with the same colours. This method again follows the taxonomy of decision tree B (Figure 13) and provides the two groups



or



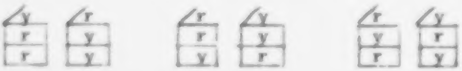
g) Grouping according to one-colour or multi-colour. One obtains two one-colour houses



and 6 multi-colour houses.

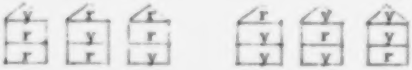
h) First continuation of g)

Subdivision of the multi-colour houses alternating in pairs



i) Second continuation of g)

Subdivision of the multi-colour houses according to the distribution of the colours, two times red - one time yellow or one time red - two times yellow

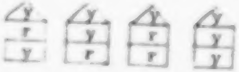


Since those continuing subdivisions c), d), e) and h), i) are not realizable at the same time, this variety of solution possibilities requires discussion within the group. Depending upon how the ships or houses look, the pupils will particularly notice different aspects; the different suggestions for a solution have to be coordinated and they have to consult about the future path. This is a typical situation in which the working group is under a strong pressure to talk, deriving from the mathematical state of affairs in its methodological development.

The pupils, who are not yet aware of the taxonomy above mentioned, but only start discovering it through their own activity, showed a behaviour that is typical in such situations. They first recite conspicuous single aspects, which at first stand next to each other in a very uncoordinated way; suspected regularities are formulated as statements which often determine the direction for the further proceedings and then prove to be right or wrong. In the fifth phase, the final coordination of the singular aspects towards a total solution has to be performed.

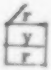
The individual analysis shows which solution strategies are favoured by the pupils:

Pupil: "Four yellow roofs"
He means



(Taxonomy f)

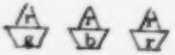
Pupil: "Wait a minute, see, and now four red roofs".
This is hypothetically formulated in analogy to the 4 houses with the yellow roofs since thus far there are only three houses with a red roof. The combination



is missing. Therefore, other pupils now react

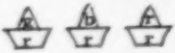
Pupil: "No, six red roofs."
Pupil: "Five red roofs."
Pupil: "There, three red roofs."
With this, this attempt to systematize remains unfinished.

Pupil: "No, three red sails first."
He means taxonomy a)



This pupil had not taken part in the discussion on the houses and now pulls the other pupils into a discussion about the boats.

Pupil: "Three red boats".



Pupil: "I'm sure there are a lot of mistakes there. Three red sails and three green sails."

Pupil: "How many blue ones? Three blue sails."
With this all the singular aspects of taxonomy a) have been formulated by the children, but still, they cannot see the total solution. The discussion on the grouping of the ships goes on. In doing so, the singular aspects are always related more closely to each other, so this can be considered to be about the transition to the fifth phase.

Fifth phase: Reaching a total solution through the combination of the singular aspects

Pupil: "Three blue ships! No, that one is red and below it's blue."

Pupil: "That's still right. Three blue ships. Should we now write three green bottom parts?"

Pupil: "Why don't you write: one green ship?"
(Change over from taxonomy a)

to taxonomy b)).

Pupil: "But this is just not possible."

Pupil: "Now I'm beginning to get it. I think that each ship is supposed to have the form red-red, blue-blue, green-green."

Pupil: "Write down, one green ship."

Pupil: "One green, one red, and one blue ship."

Pupil: "One red ship - there are more colours."

Pupil: "Six multi-colour ships." (Taxonomy b)).

Pupil: "Multi-colour - those are more colours."
Despite this objection, this attempt to solve the problem rests here after taxonomy b). A continuation according to taxonomies c), d), or e) is not tried by the pupils. They now draw their attention again to the houses and here go over to taxonomy g).

Pupil: "What now? Now we have to do the houses."

Pupil: "One yellow house, one red house."

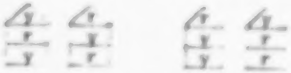
Pupil: "Six multi-coloured houses."

Here, too, the pupils are satisfied with this partial solution. They do not have the need further to refine the general overview which they now possess. In order to continue their work, it would now be necessary for the teacher to step in perhaps in the form of a new problem-posing or other guidance. One could also think of structuring the pupils' work from the very beginning towards a more differentiated work result by handing out suitable empty forms on which ships or houses have to be grouped.

The result of the work was retained by the pupils in written form

	commentary
6 yellow roofs 6 red roofs	taxonomy f)
3 red sails 3 green sails 3 blue sails	taxonomy a)
1 green ship 1 blue ship 1 red ship 6 multi-colour ships	taxonomy b)
1 yellow house 1 red house 6 multi-colour houses	taxonomy g)

The continuing subdivisions were not explicitly formulated by the children. They had, however, on the active level, silently always been using them when, for example, they placed pairs of figures - taxonomy h).



Final Remarks

The analysis of this example lets us see that group instruction also has positive effects for the identification of new insights. Obviously, the group situation favours the willingness of the children to discuss. Their different suggestions for a solution and conjectures represent a constant challenge for the spontaneous exchange of argument and counterargument. The attempts of the children at reasoning, which were often just indicated on the active level, bring light into the factual relationships, sometimes better than any verbal explanations on the part of the teacher, however well they may have been meant.

The group situation has proved successful when it comes to constructively building up a situation for solution especially in the third work phase, which dealt with the analysis of the situation which had been constructed by the pupils themselves. The joint commitment is often expressed by the pupils through statements like

- in the area of emotional concern - "Look, here, what I found."
- in the introduction of new problem-solving suggestions - "Or we can make everything yellow."
- in asking or looking for reasoning - "Why did you do that?"

On the basis of these observations, the idealism of many teachers who in their instruction classes rely on the curiosity and creative abilities of children, seems to be well-founded. Well-known theses are confirmed, and at the same time, taken out of the field of theory and are made more concrete in the school practice:

- Strong commitment and dedication on the part of the pupils with regard to the instruction can almost always be expected when the pupils are given the right to take a real part in making the decision for the directly following instruction. The participation of the pupils in the instruction gets more intensive because they have made the events in the class their own cause.
- The direct measures taken by the teacher are willingly accepted - they therefore have a positive effect on the learning progress - if they not only seem to concern the pupils' statements but also provide the pupils with a clear purpose.
- Appropriate combinational problems in favourable methodological development forms are particularly apt to be the natural basis of group discussion. The pupils can formulate solution hypotheses by themselves, attempt to verify them and give reasons for what they are doing and can independently carry out final control and control calculations.

TOPIC 19: TEACHING OF AGRICULTURAL STATISTICS

ORGANISER: DR D SINGH

TRAINING OF PERSONNEL FOR AGRICULTURAL STATISTICS IN JORDAN

Z R Takrouri, Department of Statistics, Amman

INTRODUCTION

Agricultural statistics is considered the basis of all agricultural decisions related to development plans in the agricultural sector. In most developing countries, planners try to benefit from the kinds of agricultural statistics readily available. In many cases they find that the statistics are either incorrect or incomplete. This makes them look for other sources of information available in the publications of international agencies, such as FAO, or try to compromise a figure taken from several sources.

The importance of getting fairly accurate agricultural statistics makes the programme of training the personnel of agricultural data producing agencies on better ways of improvement an urgent need. In Jordan, the government established in 1963 the Statistical Training Centre connected with the Department of Statistics. The purpose of setting up this centre was to train the governmental and non-governmental statistical personnel on how to raise the level of accuracy of the different statistical figures through learning statistical methods and their applications. A complete course (3 credit hours) in agricultural statistics was among the courses to be taught to the participants.

This paper will throw light on this kind of training of personnel, the problems faced during it, and suggestions toward improvement.

AGRICULTURAL STATISTICS

A. Definition: There is no fixed definition of agricultural statistics, but we can say that it is the statistics which deals with agriculture and covers all the agricultural aspects. But since agriculture is connected with other sectors of economy, one may be confused with some statistics whether one can consider them as agricultural statistics or not. Many examples can be given on these inter-related statistics. Agricultural constructions, food exports and imports, and loans given to farmers by credit corporations are some examples of statistics which deal with two sectors and can be considered in both sectors for specific purposes. In spite of this, one can consider any statistics which throw some information on any item related to agriculture as agricultural statistics. On the other hand, agricultural statistics covers many specific terms that ought to be defined. Usually in each of the censuses and surveys conducted in any country, some space is devoted to definitions and concepts while reporting the results. This will make the reader or data user's job much easier. A page or two of definitions in the enumerator's guide will enable him to understand the questionnaires and fill them in accurately.

B. Need for training: Agricultural statistics deals mainly with quantitative descriptions of land and livestock. The work involved is voluminous and therefore, in many countries, particularly in developing ones, sample survey technique is used for collection of data. This requires training for the persons who work on such projects. Map reading and sketch preparing need special training. For quality check purposes, accurate area measurement of selected plots may be needed. Area measurement technique has its own methods and equipment, and it is not easy for any person to learn without training. Even the conduct of crop-cutting experiments to estimate the average yield of production per area unit needs a special type of training.

From what is stated above, one can conclude that a new graduate cannot enrol himself in the agricultural statistics professional team and work with them without getting special training in the work needed in connection with agricultural statistics.

C. Surveys conducted in Jordan: So far three agricultural censuses have been conducted in Jordan since the adoption of the present constitution. A fourth one is planned to be implemented in 1983. Besides the censuses, three annual surveys dealing with land are conducted. Two of them are on a sampling basis while the third is based on a complete enumeration of lands irrigated by underground water. From these surveys, the area and production statistics of crops are obtained in the country. We are now in the process of adopting a new scheme of livestock and poultry statistics. An agricultural household survey is planned to be conducted this year to study the socio-economic aspects of rural households. This survey will provide data users with many socio-economic indicators approved by the last conference of FAO for measuring the development of rural areas. Conducting of extra specialized surveys is handicapped by the unavailability of funds and qualified personnel.

Objective methods are followed to collect the information whenever resources are made available. Area surveying for selected plots is made to correct the farmers' responses. Annual crop-cutting surveys to estimate the average yield per area unit for wheat, barley and olives have been conducted in the country since 1969.

It is obvious that the conduct of such surveys needs qualified personnel to work in the agricultural statistics section in the Jordanian Department of Statistics. This requirement has been met mainly by organising training within the country and sometimes by deputing abroad.

QUALIFICATIONS

A. Trainer: It is not an easy job to train people in how to conduct censuses and surveys especially in agriculture. At the same time, it is difficult to find qualified trainers who can be entrusted with this duty. It is difficult in the developing countries to find a qualified statistician, but it is more difficult to find a qualified agricultural statistician. Maybe this is due to the difficulty associated with the fieldwork of agricultural surveys.

A good trainer should firstly have a good academic background, preferably a first degree in agriculture. The work relating to agricultural statistics needs this background since it deals with different aspects of agriculture and this is not easy if he is not well trained in agriculture. Agricultural faculties in universities in many developing countries neglect statistics courses. This means producing good agronomists with poor statistics knowledge. This problem can be solved by studying for one or two years in a statistical institute. An experience of at least five years in agricultural statistics work should be considered an essential qualification of a good trainer. This experience will teach him how to apply the academic study in the field work. He will learn the problems that he will face and the best way to solve them. He will learn how to deal with people, how to treat them, respect them, and convince them to respond correctly. During that period of experience he will be familiar with the geographical map of the country and the location of villages, roads and administrative divisions.

B. Trainee: All the trainees should be from organizations already engaged in statistical work, or those who will be employed in this work after completing the training. Their academic study should not be less than the general secondary certificate. This will enable them to understand mathematics, sampling and English terms without much difficulty. Trainees coming from rural areas and knowing agriculture will find the training in agricultural statistics very easy.

AGRICULTURAL STATISTICS COURSE IN THE JORDANIAN STATISTICAL TRAINING CENTRE

A. Contents: It is well known that statistics are obtained mainly through field work. The volume of field work differs from one kind of statistics to another. Agricultural statistics involve a big share of field work. Thus, a satisfactory course in agricultural statistics should not ignore the role of field work in the agricultural statistics process.

The syllabus of the agricultural statistics course being taught in the Jordanian Statistical Training Centre includes the following topics:

a. Agricultural census operations.
b. Methods of collecting data.
c. Kind of sampling frames.
d. Area measurement techniques.
e. Crop-cutting experiments.
f. Farm prices.
g. Agricultural production index numbers.
h. Cost of agricultural production statistics.
i. Agricultural income accounts.
j. Preparing and reading maps and sketches.
k. Data processing methods.
l. Publication of survey reports.
m. Questionnaires.
n. Field work instructions.
o. Plan of work and budget preparing.

B. Duration: The agricultural statistics course is among the courses taught to second year participants during the second semester. The duration of this course is about 15 weeks. Three lectures of one hour's duration are given in each week. Sometimes additional time is needed in order to be able to cover the whole topics especially when some holidays take place on the day of the lectures.

Field work operations are illustrated in the class room for the first time - conducting crop-cutting as an example - then one or two lecture periods are spent at a nearby location to demonstrate them on the spot. This actual demonstration makes the operation more clear and easy to stick in mind.

The Jordanian Statistical Training Centre organised two separate agricultural courses. The first was in June 1981 for three weeks' length about agricultural censuses in cooperation with the United Nations Food and Agriculture Organization. Twenty participants from local Jordanian organisations and Arab statistical agencies attended the course. The second course was held in March 1982 for five weeks' length covering different topics in agricultural statistics in cooperation with the Arab Institute for Statistical Research and Training. Thirty participants from local and Arab statistical agencies attended the course. The evaluation forms filled by the participants in the two courses indicated that they were very useful and successful and suggested that the duration of the course should be longer.

C. Field work: Field work training is considered as complementary to training in the class room. So, whenever there is a topic in the course in agricultural statistics which needs an illustration in the field, a special time is devoted for this purpose. The field work training included the following operation:

a. Area measurement of the fields to obtain correction factors of the area reported by the respondents.

b. Crop-cutting experiments to estimate the average yield of production per area unit of the three principal crops in Jordan: wheat, barley and olives.

c. Pre-test operation to test the questionnaire and to conduct a pilot study on selected locations to examine the validity of the instructions for the whole operation, and how to obtain public cooperation.

In order to ensure the attendance of participants at the field work training, a weighting of about 40% of the course mark is devoted to it and the rest of the mark to the class examinations. Each participant is given the chance to practise by himself the field work operation by forming groups to do the work. Transportation facilities to and from the spot are provided by the Department of Statistics. Equipment required for this type of training is also provided by the Department.

PROBLEMS AND SUGGESTIONS FOR IMPROVEMENT

Training in the Jordanian Statistical Training Centre lasts for two years. The course in agricultural statistics is among the courses taught during the second year. Since that is the case, it is not easy to talk about the problems facing this course only. The following problems are faced in the course of organising the whole training programme including the agricultural statistics training:

a. The majority of the participants, especially those coming from the Gulf Countries, are poor in their academic background, while some possess very good background. This difference in their background makes the trainer's job very difficult. Such a problem can be solved by imposing strict conditions for admission of participants coming from countries having poor educational standards, or by introducing an examination for admission.

b. Many countries do not give any incentives to the successful participants, which makes them disinterested in the course.

c. In some cases the participant is not employed for statistical work after finishing his training programme. He may work with completely different work. This makes him feel disappointed toward his long training programme.

d. In Jordan there are two universities. Both do not give any recognition to study in this statistical centre. This makes the participants not eligible to complete their study in the university. Thus, many are discouraged to enrol in the training programme at the centre. A continuous exchange of views to settle the matter is going on between the administrations of the two, the universities and the Department of Statistics.

e. An obvious weakness in the English language level among participants is a serious problem. To overcome this problem, a special course in English language is given during the first year of training to all participants.

f. Due to lack of funds, the payment per one hour lecture is low. This makes the instructors look for work in other institutes or with the universities, who usually pay them better. This can be solved by increasing the governmental financial support to the Department of Statistics budget.

g. Other miscellaneous problems can be mentioned which can be solved when additional funds are available, such as the unsuitability of the lecture rooms, both in size and in location, inadequacy of library facilities, etc.

Finally, we may say that in future additional time should be allowed for more courses in agricultural statistics. The course taught at present is very crowded and can be divided into two or three specialized courses.

CONCLUSION

Agricultural statistics is very important to development plans in the agricultural sector. A successful development plan should depend on accurate figures. Getting accurate statistics requires a good training of the personnel employed for agricultural statistics in the best methods to collect, tabulate, analyze, and publish data.

The Jordanian Department of Statistics realised the importance of teaching a course on agricultural statistics to the participants enrolled in the study and training programme at the Statistical Training Centre. This course covers many topics of agricultural statistics. Also, separate training programs were organized in cooperation with some international agencies to train participants from local and Arab statistical organizations on agricultural censuses or surveys.

There are some qualifications that should be possessed by the trainer in agricultural statistics. Likewise, the trainees should possess good academic background. The availability of these qualifications will lead to the best results.

Since the field work is a very important element in agricultural statistics, it is essential that the contents of the course should contain both theoretical lectures and practical field work. The minimum duration of such a course should not be less than 15 weeks.

A few problems that are being faced by the Statistical Training Centre in organising the course can be solved by suitable action and by making available additional fund resources.

ISBN 0 946554 00 5 (set of two volumes)

ISBN 0 946554 01 3 (Volume I)

ISBN 0 946554 02 1 (Volume II)

Designed and printed by the University of Sheffield Printing Unit

END

DEPT. OF EDUCATION

NAT'L INSTITUTE OF EDUCATION

ERIC

DATE FILMED

MARCH

17 1986